The international journal of science / 30 March 2023

outline
Acute kidney injury

nature



SWAHILI ANCESTRY

Medieval DNA reveals African and Asian mixing along East Africa's coast

Seeking relief
The complex barriers to treatments for chronic pain

Balanced approach
How to bolster research integrity by addressing inequity

Particle probed
Measurement of proton's mass radius finds small dense core

VISIT…

LANZAROTE Caliente.com

Nature.2023.04.01

[Sat, 01 Apr 2023]

- **This Week**
- **News in Focus**
- **Books & Arts**
- **Opinion**
- **Work**
- **Research**
- **Amendments & Corrections**
- **Nature Outline**

This Week

- **Treat pain as a priority, not an afterthought** [29 Mar 2023]
  • Chronic pain is real and ruins lives — medical attitudes to it must change.
- **Global action on water: less rhetoric and more science** [28 Mar 2023]
  • The UN’s first water conference in decades put the spotlight on a vital and troubled resource. But to widen access and resolve disputes, one thing is needed above all: data.

- EDITORIAL
- 29 March 2023

Treat pain as a priority, not an afterthought

Chronic pain is real and ruins lives — medical attitudes to it must change.



Chronic pain has many manifestations and numerous causes.Credit: Salwan Georges/The Washington Post/Getty

Why does it hurt?

This seemingly simple question doesn't always have a simple answer. Pain is generally regarded by researchers and medical practitioners as the body's warning system for a biomechanical problem: it is the alarm for when a

finger has a cut, a joint is inflamed or a tumour is growing. Medical professionals will treat the underlying problem, and the pain is expected to recede. But what happens when the alarm doesn't turn off? What happens when the pain doesn't stop, long after tissues have healed? What happens when there's pain even in the absence of tissue damage? Some answers can be found in a Feature in *Nature* this week.



Chronic pain can be treated — so why are millions still suffering?

Pain itself is a medical condition. Chronic pain, generally defined as pain that lasts for more than three months, affects millions of people worldwide every day. It is likely to affect more of us in the coming decades, in line with increases in the incidence of obesity, diabetes and autoimmune diseases. All of these conditions increase a person's likelihood of developing chronic pain. Pain is also one of the symptoms that can linger after a SARS-CoV-2 infection. Even in triple-vaccinated adults, around 4% of people with breakthrough infections of the Omicron variant develop long COVID.

Chronic pain has many manifestations and numerous causes, and basic research in neuroscience and immunology has contributed much to our understanding of it. Geneticists and epidemiologists are probing the heritable and environmental factors that contribute to the risks. It is also becoming clearer that chronic pain is tied to a complex mix of neurological,

immunological, psychological and social factors. Various factors and processes can drive a person's pain and might change over time.

Wanted: integrative care

But the multifactorial nature of pain is often not recognized by the medical establishment, especially by clinicians. Medical research is starting to engage with this biopsychosocial model of pain, and therapies have been developed that can — to some extent at least — help to quieten those unrelenting alarms. But accessing the right combination of treatments, or any treatment at all in some cases, is not easy. One considerable hurdle is the lack of recognition of the value of interventions such as yoga, acupuncture and psychotherapy, in addition to medicines. In some cases, health-care policies and insurance systems are also a barrier.



Long COVID: let patients help define long-lasting COVID symptoms

For example, integrative care (such as that involving pain psychologists, physical therapists and spinal specialists) can sometimes be more effective in treating chronic pain than are single interventions, such as steroid injections. However, in the United States, health-insurance plans often exclude these more-complex treatment plans. Furthermore, some US health-care providers are known to favour those interventions for which they are paid more by insurers, even if there are more effective (but less lucrative) alternatives. If a clinician is paid the same for administering a 30-minute

steroid injection as for providing a much longer course of treatment, this could be seen as incentivizing the shorter, more lucrative option. The country's lack of both universal health coverage and mandatory paid sick leave also prevents many people from benefiting from the treatment options available to them.

Change societal attitudes

Of course, problems surrounding access to treatments are not unique to countries without universal health care. In Australia, it's not unusual for people to have to drive for five or six hours to reach a pain clinic. In the United Kingdom, people using the National Health Service typically face a year-long waiting list before they can start a pain-management programme.

Many of the issues surrounding access to treatments can be traced to individual and societal attitudes that fail to prioritize chronic pain and chronic conditions. Some even fail to recognize that a person's pain is genuine. Despite a wealth of evidence for the neurological underpinnings of fibromyalgia, for example, a study last year revealed that some UK doctors stated in a survey that they don't think it is a distinct, or real, condition (N. Wilson *et al. BMC Health Serv. Res.* **22**, 989; 2022). Some of the doctors' responses were so offensive to patients that they were not published, say the study's authors.

Researchers are clearly helping to answer the question 'why does it hurt?', but chronic-pain research needs to be more appropriately funded. The US National Institutes of Health includes pain-research funding in the same category as addiction research, in part owing to the complicated legacy of the use of opioid drugs to treat pain. Both subjects are important and there's no problem in classifying the funding in this way, in principle. However, two-thirds of one of the world's largest funding programmes for pain and addiction, the US HEAL Initiative, goes to addiction studies, and one-third to those researching pain. That logic is more difficult to parse.

Approaches that fail to prioritize the many complex drivers of chronic pain and even deny its existence are causing unnecessary suffering for millions. Reform is needed to ensure that people experiencing chronic pain are able to

access treatments that are most likely to help them. And the stigma and bias that characterize the medical and societal narrative around chronic pain must end.

Nature **615**, 765 (2023)

doi: https://doi.org/10.1038/d41586-023-00885-6

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00885-6

- EDITORIAL
- 28 March 2023

Global action on water: less rhetoric and more science

The UN’s first water conference in decades put the spotlight on a vital and troubled resource. But to widen access and resolve disputes, one thing is needed above all: data.



Baghdad on the Tigris: this week Iraq became the first country in the Middle East to join the UN water convention.Credit: Getty

Forty-six years. That’s how long it took the United Nations to organize a high-level conference on water. The 3-day event, held last week in New York, was brief compared with its 12-day predecessor, which took place in Mar del Plata, Argentina, in 1977. That meeting helped to catalyse the drafting of the UN’s 1992 water convention, an international agreement through which countries agree to cooperate on the use and protection of

shared waterways. Delegates to this month's conference did not make any binding commitments, but countries are converging on several ideas. One is to establish a panel of scientists to regularly advise on water issues. Another is for the UN to appoint a 'water envoy', a high-level diplomat representing secretary-general António Guterres. The conference also increased the visibility of the 1992 convention.

Each of these is an important step. The appointment of a UN envoy would send a powerful signal that this is a hot topic, high on the agenda of world leaders, and that a boost to the science is overdue. For decades, there's been much rhetoric around applying science to water resources at the intergovernmental level, but action has been piecemeal. The 1977 conference report (known as the Mar del Plata Action Plan) and the 1992 convention recognize the need for research and development. This need was most recently incorporated into the UN's Sustainable Development Goals (SDGs) in 2015; SDG6 aims to achieve equitable access to safe water and sanitation by 2030. These agreements also mention the need for countries to share data — for example, on river flows, weather and climate, and water quality. But with some two billion people still lacking safe water in their homes, SDG6 is a long way from being achieved.



The world faces a water crisis — 4 powerful charts show how

The necessity of studying and sharing such data is underscored by a Comment in *Nature* written by climate scientist Johan Rockström, economist Mariana Mazzucato and their colleagues. Among other things, they describe how land-management practices in one country can impact atmospheric water flows in other countries. But to properly take advantage of such knowledge, countries need to both collect and share data.

One option proposed by the UN's science agency UNESCO is modelled on the Intergovernmental Panel for Climate Change, and would produce periodic global literature reviews to be signed off by governments. Such an idea needs rigorous analysis and testing. Separately, there is also a real need for input from independent scientists on water disputes between nations.

At the start of the year, Mohammed Basheer, a water-resources economist at the Humboldt University of Berlin, and his colleagues published a modelling study addressing a protracted dispute over Africa's largest hydroelectric dam on the Nile River. The work showed how Egypt, Ethiopia and Sudan, the three countries involved, could benefit if each was willing to compromise (M. Basheer *et al*. *Nature Clim. Change* **13**, 48–57; 2023). However, it's not clear whether these recommendations will find their way into negotiations. Relations between Egypt and Ethiopia are fraught, and there is no clear path for the provision of independent scientific advice on the dam.



Why we need a new economics of water as a common good

A study by Patience Mukuyu at the International Water Management Institute in Pretoria, South Africa, and her colleagues, published in early March, reaffirms that relatively few countries that share watercourses are cooperating, such as by sharing data (P. Mukuyu *et al. Water Int.* https://doi.org/j36p; 2023). Co-author Alistair Rieu-Clarke at Northumbria University, UK, who studies transboundary water issues, emphasizes that there are demonstrable benefits to data sharing. In September 2021, four African countries — Gambia, Guinea Bissau, Mauritania and Senegal — agreed to cooperate (including by sharing data) on a giant underground aquifer, which covers more than 300,000 square kilometres and supplies water to 80% of their combined populations.

The list of transboundary water disputes is lengthening. The Indus Waters Treaty, a 63-year-old agreement between India and Pakistan on sharing the Indus River's waters, is in trouble. The melting of glaciers as a result of climate change is affecting river flows; at the same time, both countries have plans in the works to build dams. In January, India announced that it wants to renegotiate the treaty to take account of a changed environmental and geopolitical situation, rather than work within what it sees as an outdated framework. However, Pakistan wants to continue to resolve disagreements using the treaty's dispute-settlement procedures. Elsewhere, Iran, Iraq, Syria and Turkey have long been at odds over the shared Euphrates–Tigris river basin, where water availability will also be affected by climate change.

The idea of a UN water envoy and science panel offer an opportunity to make a difference to such disputes. More countries that share water resources also need to sign and ratify the 1992 UN water convention. Last week, Iraq, Nigeria and Panama were among ten countries to either do so or announce that they intend to do as much.

The UN and its member states have made a positive start. Increased visibility of the 1992 convention will pay dividends, as will boosting research to help policymakers to better understand and resolve water crises, although the devil will be in the detail. Science, as Basheer told *Nature* earlier in the year, can show countries how to "help each other, look after each other and look out for each other", when water disputes are being discussed. Ultimately, countries need to accept that scientists must have a seat at the table during these discussions.

Nature **615**, 766 (2023)

doi: https://doi.org/10.1038/d41586-023-00886-5

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00886-5

| Section menu | Main menu |

News in Focus

- **COVID-origins study links raccoon dogs to Wuhan market: what scientists think** [21 Mar 2023]
  • Some say the analysis supports the hypothesis that the virus that causes COVID-19 spilled over from an animal — but falls short of definitive proof.
- **GPT-4 is here: what scientists think** [16 Mar 2023]
  • Researchers are excited about the AI — but many are frustrated that its underlying engineering is cloaked in secrecy.
- **The world faces a water crisis — 4 powerful charts show how** [21 Mar 2023]
  • Hundreds of millions of people lack access to safe water and sanitation. Will the first UN conference on water in nearly 50 years make a difference?
- **How stem cells make a human brain** [16 Mar 2023]
  • Technical 'tour de force' allows researchers to trace the family tree of crucial brain cells.
- **Abel Prize: pioneer of 'smooth' physics wins top maths award** [22 Mar 2023]
  • Luis Caffarelli's work includes equations underpinning physical phenomena, such as melting ice and flowing liquids.
- **'Open for business': risk-taking US health agency ready to spend $2.5-billion budget** [21 Mar 2023]
  • One year after the launch of ARPA-H, Nature talks to director Renee Wegrzyn about her vision for the agency.
- **Strife at eLife: inside a journal's quest to upend science publishing** [17 Mar 2023]
  • Editors threaten to resign over 'no-reject' model that others see as the future of research journals.
- **Chronic pain can be treated — so why are millions still suffering?** [28 Mar 2023]
  • Science is beginning to uncover the multiple processes driving persistent pain. But connecting people with treatments that will help them remains a challenge.

- NEWS
- 21 March 2023
- Correction 22 March 2023

COVID-origins study links raccoon dogs to Wuhan market: what scientists think

Some say the analysis supports the hypothesis that the virus that causes COVID-19 spilled over from an animal — but falls short of definitive proof.

- Smriti Mallapaty



Two raccoon dogs in a Chinese market. DNA from this species was found in positive SARS-CoV-2 swabs taken from the Huanan market, from where

COVID-19 is thought to have spread.Credit: Greg Baker/AFP via Getty

Raccoon dogs, bamboo rats, palm civets: these are just some of the animals whose DNA has been found in swabs taken from the Huanan Seafood Wholesale Market in Wuhan, China, which has been linked to the origin of the COVID-19 pandemic. The swabs also tested positive for SARS-CoV-2, which causes the disease. The analysis — posted on 20 March to the research repository Zenodo[1] — provides evidence supporting the hypothesis that SARS-CoV-2 spilled over from animals to humans at the market, say some researchers.

"Of course, this is not direct evidence," says Leo Poon, a virologist at the University of Hong Kong, "but this is the best we can get now, because all the animals have been eliminated from the market and we don't have swabs of the animals."

On 16 March, *The Atlantic* first reported on the analysis. The Zenodo posting, which has not been peer reviewed, is the first time that the full work has been released publicly — which could pave the way for follow-up studies, such as an investigation of where the animals in the market came from that could help to pinpoint the pandemic's start.

The swabs were collected in early 2020, after the market was shut down and cleared of animal products.

Particularly notable was the raccoon-dog mitochondrial DNA found in six samples from two stalls. These small fox-like animals are susceptible to SARS-CoV-2, and can spread the infection to other raccoon dogs without showing clear signs of sickness[2]. Raccoon dogs and masked palm civets have also been found with infections of viruses that are almost identical to the one that causes severe acute respiratory syndrome, which is related to SARS-CoV-2 and caused an outbreak in people in 2003. And work in palm-civet cells indicates that the creatures could possibly become infected with SARS-CoV-2[3].

The researchers were looking for evidence of mammals, which could have been intermediate hosts of the virus. They identified near-complete mitochondrial-DNA sequences — each some 16,000 base pairs long — for

five species, including raccoon dog (*Nyctereutes procyonoides*), Malayan porcupine (*Hystrix brachyura*), Amur hedgehog (*Erinaceus amurensis*), masked palm civet (*Paguma larvata*) and hoary bamboo rat (*Rhizomys pruinosus*). “It’s remarkable to have a list,” says study co-author Alex Crits-Christoph, who is a computational biologist at a non-profit organization and is based in Baltimore, Maryland.

Studies suggest that a progenitor of SARS-CoV-2 probably originated in bats, but how it passed from bats to people is not clear. Scientists suspect an intermediate host was involved, but they have not found strong evidence, which has led to speculation that the virus could have instead leaked — deliberately or accidentally — from a laboratory in Wuhan.

The new study does not confirm whether the animals themselves were infected with the virus. But establishing that the animals were at the market suggests that the pandemic could have had an animal origin, say some researchers.

Others are more cautious about drawing such conclusions from the analysis. The work confirms which animal species were at the market, but because it includes only samples that had tested positive, “there is no data in this work associating SARS-CoV-2 with the presence of any of these animals”, says Justin Kinney, a quantitative biologist at Cold Spring Harbor Laboratory in New York.

“These data do not provide a definitive answer to the question of how the pandemic began,” said Tedros Adhanom Ghebreyesus, director-general of the World Health Organization (WHO), at a press event on 17 March, after he had been briefed in advance of the report’s release. “But every piece of data is important in moving us closer to that answer,” he said.

Appearance and disappearance

The swabs were first mentioned in a preprint paper by George Gao[4], former director of the Chinese Center for Disease Control (China CDC), and colleagues, posted on Research Square in February 2022. Researchers have

asked for those data to be made public on several occasions, without success.

But Florence Débarre, an evolutionary biologist at the French National Centre for Scientific Research, who is a co-author of the latest report, discovered the data almost by chance on the public data repository GISAID on 4 March. “Basically, [they’re] the ones we’ve been waiting for, for a year,” she says.

In total, Débarre and her colleagues downloaded around half a terabyte of genomic data from some 50 samples. The data included almost all the swabs taken from drains, market stalls and the ground between January and February 2020 that had tested positive for SARS-CoV-2. “We’re talking about millions and millions of sequencing reads,” says Crits-Christoph.

But soon after the researchers downloaded the data, they disappeared from public view on GISAID.

In an e-mail to *Nature*, a GISAID spokesperson said it does not delete records, but that contributors often update their records, which renders them “temporarily invisible”. The market-swab data “are currently being updated with newer and additional data as part of a manuscript currently under review”. (The Gao paper is marked as ‘under review’ at Nature Portfolio. *Nature*’s news team is independent of its publisher Springer Nature, which publishes Nature Portfolio journals.)

Débarre and her colleagues say they reached out to authors of the preprint and to the China CDC to collaborate on the analysis, but the China CDC declined. On 14 March, her colleagues, as well as members of the China CDC, presented their work to the WHO at a meeting of the Scientific Advisory Group for the Origins of Novel Pathogens, a body tasked with investigating outbreaks, including the origins of SARS-CoV-2.

At the WHO press briefing on 17 March, Tedros called on the China CDC researchers to be transparent in sharing data. “These data could have and should have been shared three years ago,” he said.

Gao did not respond to *Nature*’s requests for comment.

Débarre recognizes that it has taken a long time to share this data, but acknowledges the environmental sampling and sequencing work conducted by the China CDC. “They did what needed to be done,” she says. “We wouldn’t have this data, if it wasn’t for their work.”

Wild match

Previous studies have found that Chinese bamboo rats (*Rhizomys sinensis*) and raccoon dogs were sold at the Huanan market between May 2017 and November 2019[5], and that several stalls were probably selling live or freshly killed mammals or other meat products[6]. The data confirm that “exactly what we thought was there in the market, was there in the market”, says study co-author Joel Wertheim, a molecular epidemiologist at the University of California, San Diego.

Most of the SARS-CoV-2-positive samples and most of the samples containing wild-animal DNA were found in the same area, which Poon says “agrees with the hypothesis that maybe the virus came from animals”. Many positive swabs from the eastern section of the market contained human DNA — suggesting that there was some human-to-human transmission at the market in early 2020, says Poon.

In one sample from a cart used to move cages, the researchers found material that matched the raccoon-dog genome, and very little that matched the human genome. “The most logical hypothesis is that raccoon dogs were infected by SARS-CoV-2 and shed the virus,” says Poon.

But it is possible that the site was contaminated with SARS-CoV-2 RNA from other sources. “There is no evidence that the virus came from an infected raccoon dog; it could just as easily have come from an infected person,” says Kinney.

More data needed

Many more forensic insights could come from the data if they were made public again, says Crits-Christoph. For example, by looking at the ratio of RNA to DNA for different animal species, researchers could determine how

recently those animals were at the market, because DNA typically persists in the environment longer than RNA does. But more details are needed about how the samples were collected, processed and sequenced, says Wertheim.

A close study of the RNA data could possibly determine whether the animals at the market were sick, adds Crits-Christoph. Does it seem as though the creatures were in “a disease state, or a healthy state?” he asks. “There’s all sorts of wonderful, complex questions in the data.”

There are more data that have not yet been shared — including environmental swabs that were negative for SARS-CoV-2. Those data are important to analyse, to get a better picture of what was happening at the market, says Wertheim.

Nature **615**, 771-772 (2023)

doi: https://doi.org/10.1038/d41586-023-00827-2

Updates & Corrections

- **Correction 22 March 2023**: This article originally stated that Débarre presented the group’s findings to the WHO. It was her colleagues Michael Worobey, Edward Holmes and Kristian Andersen. The text has been corrected.

References

1. Crits-Christoph, A. *et al.* Preprint at Zenodo https://doi.org/10.5281/zenodo.7754299 (2023).

2. Freuling, C. M. *et al. Emerg. Infect. Dis.* **26**, 2982–2985 (2020).

3. Meng, L. *et al. ISME J.* **17**, 549–560 (2023).

4. Gao, G. *et al.* Preprint at Research Square https://doi.org/10.21203/rs.3.rs-1370392/v1 (2022).

5. Xiao, X. *et al. Sci. Rep.* **11**, 11898 (2021).

6. Worobey, M. *et al. Science* **377**, 951–959 (2022).

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00827-2

- NEWS
- 16 March 2023

GPT-4 is here: what scientists think

Researchers are excited about the AI — but many are frustrated that its underlying engineering is cloaked in secrecy.

- Katharine Sanderson



The GPT-4 artificial-intelligence model is not yet widely available.Credit: Jaap Arriens/NurPhoto via Getty Images

Artificial intelligence company OpenAI this week unveiled GPT-4, the latest incarnation of the large language model that powers its popular chatbot ChatGPT. The company says GPT-4 contains big improvements — it has

already stunned people with its ability to create human-like text and generate images and computer code from almost any prompt. Researchers say these abilities have the potential to transform science — but some are frustrated that they cannot yet access the technology, its underlying code or information on how it was trained. That raises concern about the technology's safety and makes it less useful for research, say scientists.

GPT-4 was released on 14 March, and one upgrade is that it can now handle images as well as text. And as a demonstration of its language prowess, OpenAI, which is based in San Francisco, California, says that it passed the US bar legal exam with results in the ninetieth centile, compared with the tenth centile for the previous version of ChatGPT. But the technology is not yet widely accessible — only paying subscribers to ChatGPT so far have access.



ChatGPT listed as author on research papers: many scientists disapprove

"There's a waiting list at the moment so you cannot use it right now," says Evi-Anne van Dis, a psychologist at the University of Amsterdam Medical Centers. But she has seen demos of GPT-4. "We watched some videos in which they demonstrated capacities and it's mind-blowing," she says. One instance, she recounts, was a hand-drawn doodle of a website, which GPT-4 used to produce the computer code needed to build that website, as a demonstration of the ability to handle images as inputs.

But there is frustration in the science community over OpenAI's secrecy around how the model was trained and what data were used, and how GPT-4 actually works. "All of these closed-source models, they are essentially dead ends in science," says Sasha Luccioni, a research scientist specializing in climate at HuggingFace, an open-source AI cooperative. "They [OpenAI] can keep building upon their research, but for the community at large, it's a dead end."

'Red team' testing

Andrew White, a chemical engineer at the University of Rochester, New York, has had privileged access to GPT-4 as a 'red-teamer': a person paid by OpenAI to test the platform to try and make it do something bad. He has had access to GPT-4 for the past six months, he says. "Early on in the process, it didn't seem that different," compared with previous iterations.



[Abstracts written by ChatGPT fool scientists]

He put to the bot queries about what chemical reaction steps were needed to make a compound, predict the reaction yield and choose a catalyst. "At first, I was actually not that impressed," White says. "It was really surprising because it would look so realistic, but it would hallucinate an atom here. It would skip a step there," he adds. But when, as part of his red-team work, he

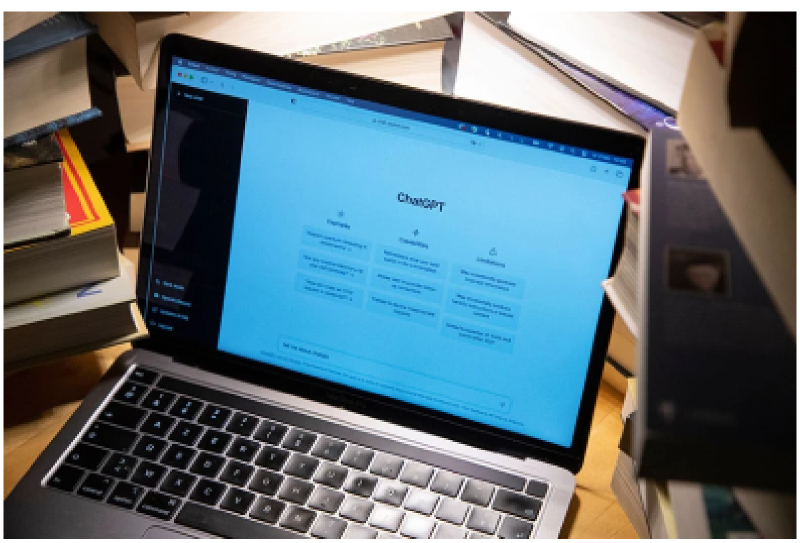
gave GPT-4 access to scientific papers, things changed drastically. “It made us realize that these models maybe aren’t so great just alone. But when you start connecting them to the Internet to tools like a retrosynthesis planner, or a calculator, all of a sudden, new kinds of abilities emerge.”

And with those abilities come concerns. For instance, could GPT-4 allow dangerous chemicals to be made? With input from people such as White, OpenAI engineers fed back into their model to discourage GPT-4 from creating dangerous, illegal or damaging content, White says.

Fake facts

Outputting false information is another problem. Luccioni says that models such as GPT-4, which exist to predict the next word in a sentence, can’t be cured of coming up with fake facts — known as hallucinating. “You can’t rely on these kinds of models because there’s so much hallucination,” she says. And this remains a concern in the latest version, she says, although OpenAI says that it has improved safety in GPT-4.

Without access to the data used for training, OpenAI’s assurances about safety fall short for Luccioni. “You don’t know what the data is. So you can’t improve it. I mean, it’s just completely impossible to do science with a model like this,” she says.



[How Nature readers are using ChatGPT]

The mystery about how GPT-4 was trained is also a concern for van Dis's colleague at Amsterdam, psychologist Claudi Bockting. "It's very hard as a human being to be accountable for something that you cannot oversee," she says. "One of the concerns is they could be far more biased than, for instance, the bias that human beings have by themselves." Without being able to access the code behind GPT-4, it is impossible to see where the bias might have originated, or to remedy it, Luccioni explains.

Ethics discussions

Bockting and van Dis are also concerned that these AI systems are increasingly owned by big tech companies. The researchers want to make sure the technology is properly tested and verified by scientists. "This is also an opportunity because collaboration with big tech can, of course, speed up processes," she adds.

Van Dis, Bockting and colleagues argued earlier this year that there is an urgent need to develop a set of 'living' guidelines to govern how AI and tools such as GPT-4 are used and developed. They are concerned that any legislation around AI technologies will struggle to keep up with the pace of development. Bockting and van Dis have convened a summit of invited participants at the University of Amsterdam on 11 April to discuss these concerns, with representatives from organizations including the science-ethics committee of UNESCO, the United Nations' scientific and cultural agency, the Organisation for Economic Co-operation and Development and the World Economic Forum.

Despite the concern, GPT-4 and its future iterations will shake up science, says White. "I think it's actually going to be a huge infrastructure change in science, almost like the Internet was a big change," he says. It won't replace scientists, he adds, but could help with some tasks. "I think we're going to start realizing we can connect papers, data programs, libraries that we use and computational work or even robotic experiments."

Nature **615**, 773 (2023)

doi: https://doi.org/10.1038/d41586-023-00816-5

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00816-5

- NEWS EXPLAINER
- 21 March 2023
- Correction 22 March 2023

The world faces a water crisis — 4 powerful charts show how

Hundreds of millions of people lack access to safe water and sanitation. Will the first UN conference on water in nearly 50 years make a difference?

- Miryam Naddaf



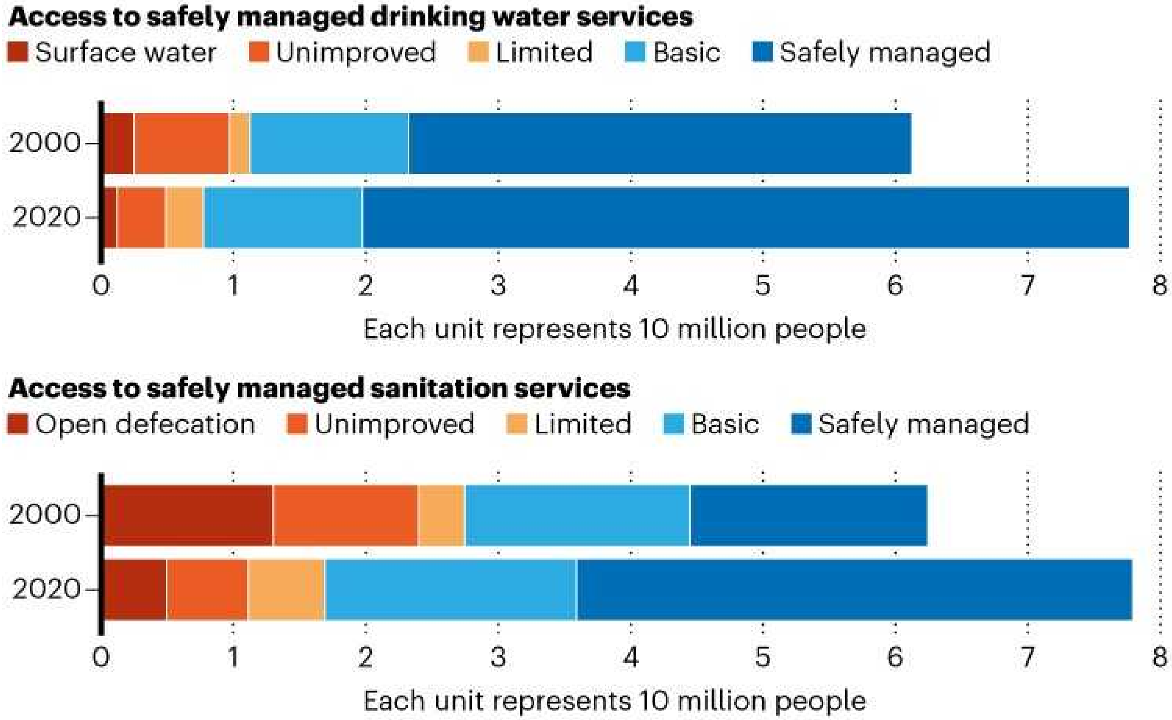
As many as 43,000 people might have died as a result of drought in Somalia last year, according to a report published this week by the Somalian

government, the UN children’s agency UNICEF and the World Health Organization.Credit: Jerome Delay/AP/Shutterstock

The United Nations water conference starts tomorrow. Co-hosted by the Netherlands and Tajikistan, the three-day event will take place at UN headquarters in New York and will be the first such event in nearly half a century. During that time, a rising number of people around the world have gained access to safe water and sanitation (see ‘A tale of two halves’) — except in sub-Saharan Africa (see ‘The neglect of Africa’), where the numbers without safe drinking-water services are greater than they were in 2000. Globally, around 500 million people are compelled to use open defecation, and millions more rely on contaminated water supplies. Can this conference make a difference?

A TALE OF TWO HALVES

Between 2000 and 2020, more people gained access to safe water and sanitation. But nearly 500 million people are still using open defecation.



Access to safely managed drinking water services. **Surface water:** Drinking water directly from a river, dam, lake, pond, stream, canal or irrigation canal; **Unimproved:** Drinking water from an unprotected dug well or unprotected spring; **Limited**: Drinking water from an improved source, for which collection time exceeds 30 minutes for a round trip, including queuing; **Basic:** Drinking water from an improved source, provided collection time is not more than 30 minutes for a round trip, including queuing; **Safely managed:** Drinking water from an improved source that is accessible on premises, available when needed and free from faecal and priority chemical contamination.

Access to safely managed sanitation services. **Open defecation:** Disposal of human faeces in fields, forests, bushes, open bodies of water, beaches or other open places, or with solid waste; **Unimproved:** Use of pit latrines without a slab or platform, hanging latrines or bucket latrines; **Limited:** Use of improved facilities that are shared with other households; **Basic:** Use of improved facilities that are not shared with other households; **Safely managed:** Use of improved facilities that are not shared with other households and where excreta are safely disposed of in situ or removed and treated off-site.

©nature

Source: WHO/UNICEF

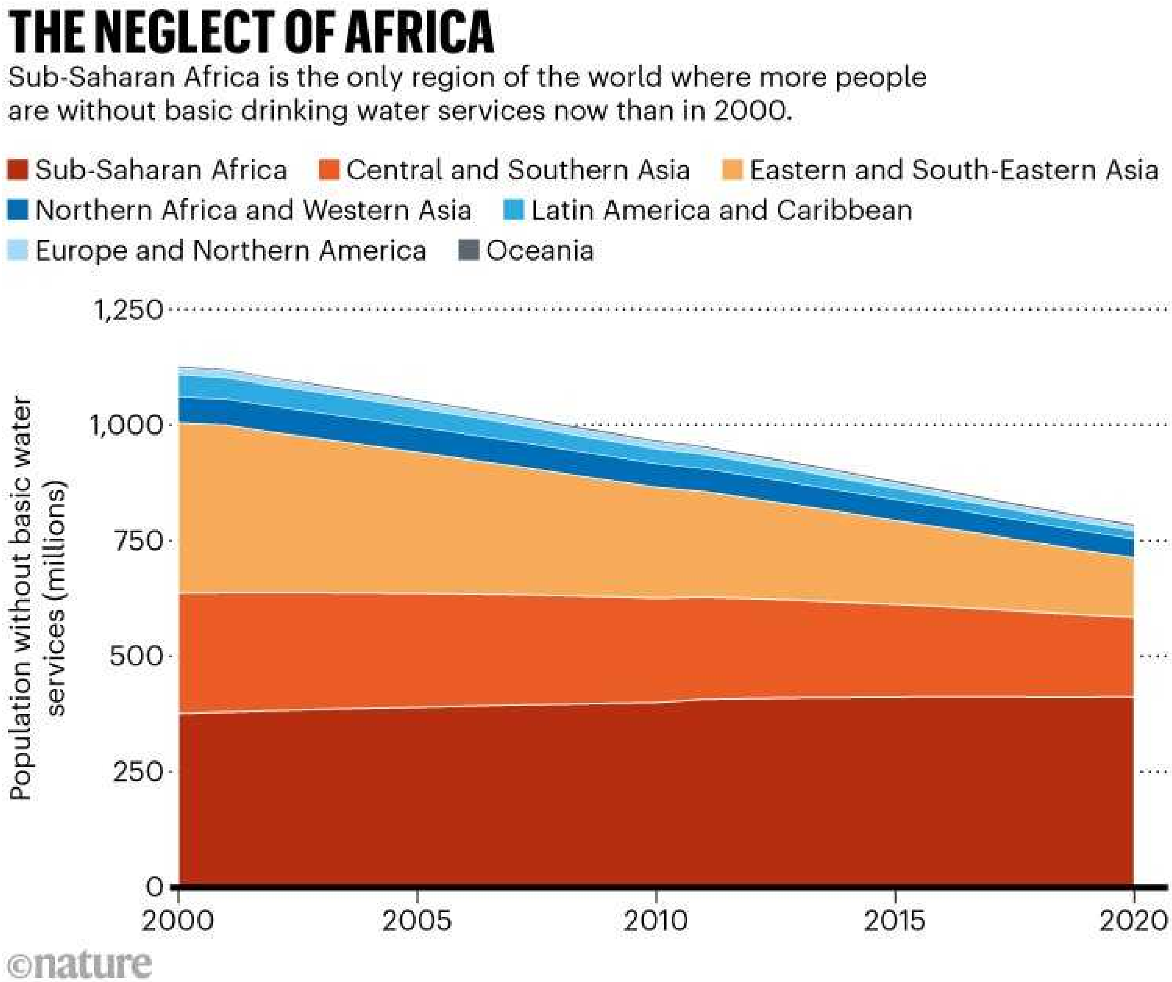
Why has it taken the United Nations 46 years to organize a conference dedicated to water?

The simplest answer is that water (as a standalone topic) has not been high on the international sustainable-development policy agenda — at least, not until now, says Rachael McDonnell, deputy director-general for research for development at the International Water Management Institute, based in Rome.

The first UN water conference took place in Mar del Plata, Argentina, in 1977. Representatives of 118 countries and territories met over 12 days and issued the Mar del Plata Action Plan, which recommended that countries achieve universal clean water and sanitation by 1990 to avoid a global water crisis by the end of the twentieth century.

Several low-income countries asked for financial support, but were rebuffed, and instead a study was proposed on how to finance water projects, as *Nature* reported at the time.

In 2015, the international community set a 2030 target (under the UN Sustainable Development Goals, SDGs) for providing clean water and sanitation to all. As of 2020, some 2 billion people still lacked safe drinking water in their homes, and around one-third of people did not have basic handwashing facilities at home, according to data from the World Health Organization (WHO) and the UN children's agency UNICEF. At current rates of improvement, 1.6 billion people will still lack safe drinking water at home by the 2030 deadline.



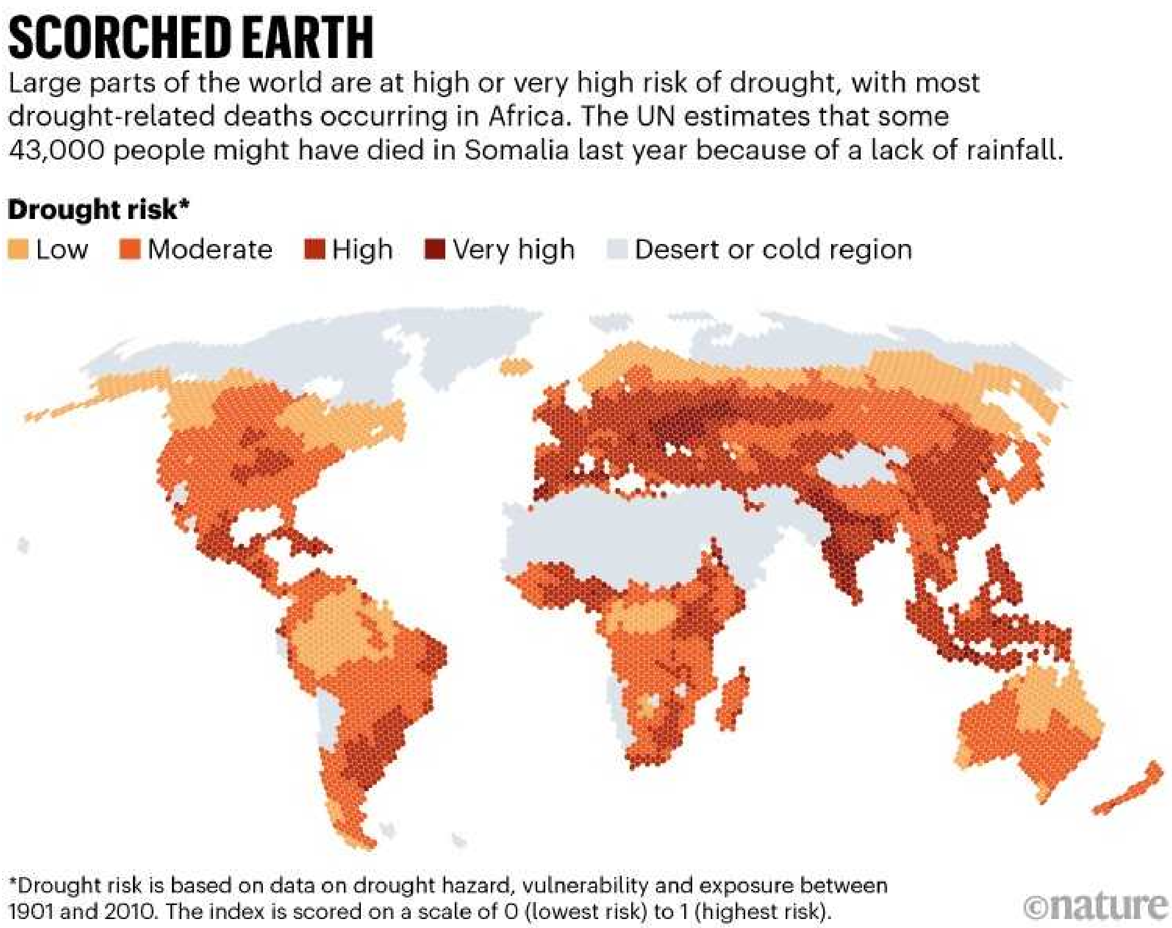
Source: WHO/UNICEF

Water was not a priority for the September 2021 UN conference on food security, nor last year's climate COP27 in Egypt, says Henk Ovink, the Netherlands' special envoy for international water affairs. Water must have a firm place in any follow-up UN process, he stresses. These include the UN Food Systems Stocktaking Moment that will take place in Rome in July, the SDG Summit in September in New York, and COP28 in Dubai in November. "We can't wait another 46 years because what is happening is just too awful at the moment, and it's going to get worse," says McDonnell.

Where is the crisis at its worst?

The water crisis is worst in low-income countries — for example, an estimated 70% of the population of sub-Saharan Africa lacks safe drinking-water services.

The conference needs to prioritize addressing water insecurity in vulnerable communities and those in conflict and post-conflict settings, says Carol Cherfane, director of the Arab Centre for Climate Change Policies, a think-tank connected to the UN based in Beirut. A report published this week by UNICEF and the WHO says that as many as 43,000 people might have died last year from drought in Somalia (see 'Scorched Earth').



Source: IPCC

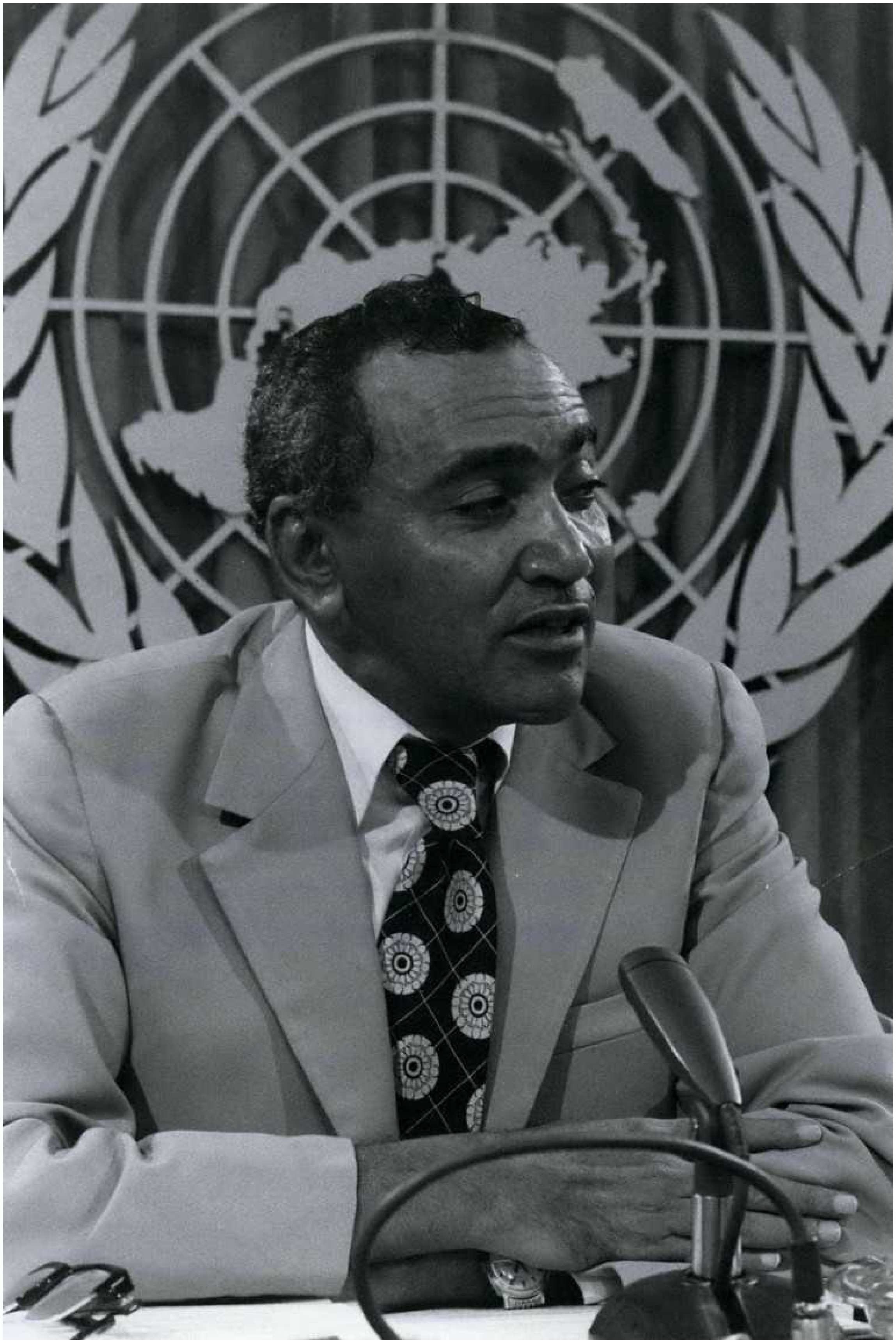
Providing better access to water for health care and sanitation is another urgent priority. Too many people have no choice but to use contaminated water supplies (see 'Faecal contamination'). In 2021, one in 10 health-care

facilities worldwide had no sanitation services and some 857 million people had no water service at their health-care facility, according to an earlier joint report also by the WHO and UNICEF summarizing 20 years of data on water and sanitation.

The conference will also discuss a plan for countries that share their water resources to communicate more effectively. The Transboundary Water Cooperation Coalition was launched at the Paris headquarters of the UN science agency UNESCO at the end of last year. This will be particularly important for countries in the Middle East and North Africa. Around two-thirds "of the water resources in Arab states flow from outside of their national borders", says Cherfane. "A transboundary basin-level approach is very important, not as an instrument of creating conflicts, but as an instrument of creating opportunity for conversations and cooperation and coordination."

How is climate change affecting water resources?

Around half of the world's population is already at risk of severe water scarcity for at least some of the year, according to the latest (sixth) assessment report from the Intergovernmental Panel on Climate Change, published this week. This number is likely to increase owing to the effects of climate change, such as heavy precipitation, flooding, drought and wildfire events. If global temperatures reach 1.5°C above pre-industrial temperatures, extreme agricultural (soil moisture) drought is expected to be twice as likely in many parts of the world.



Yahia Abdel Mageed, Sudan’s then-irrigation minister, was secretary-general of the first UN water conference in 1977. Calls at that meeting for financial help for low-income countries were rebuffed.Credit: Keystone Pictures USA/Zuma Press/Alamy

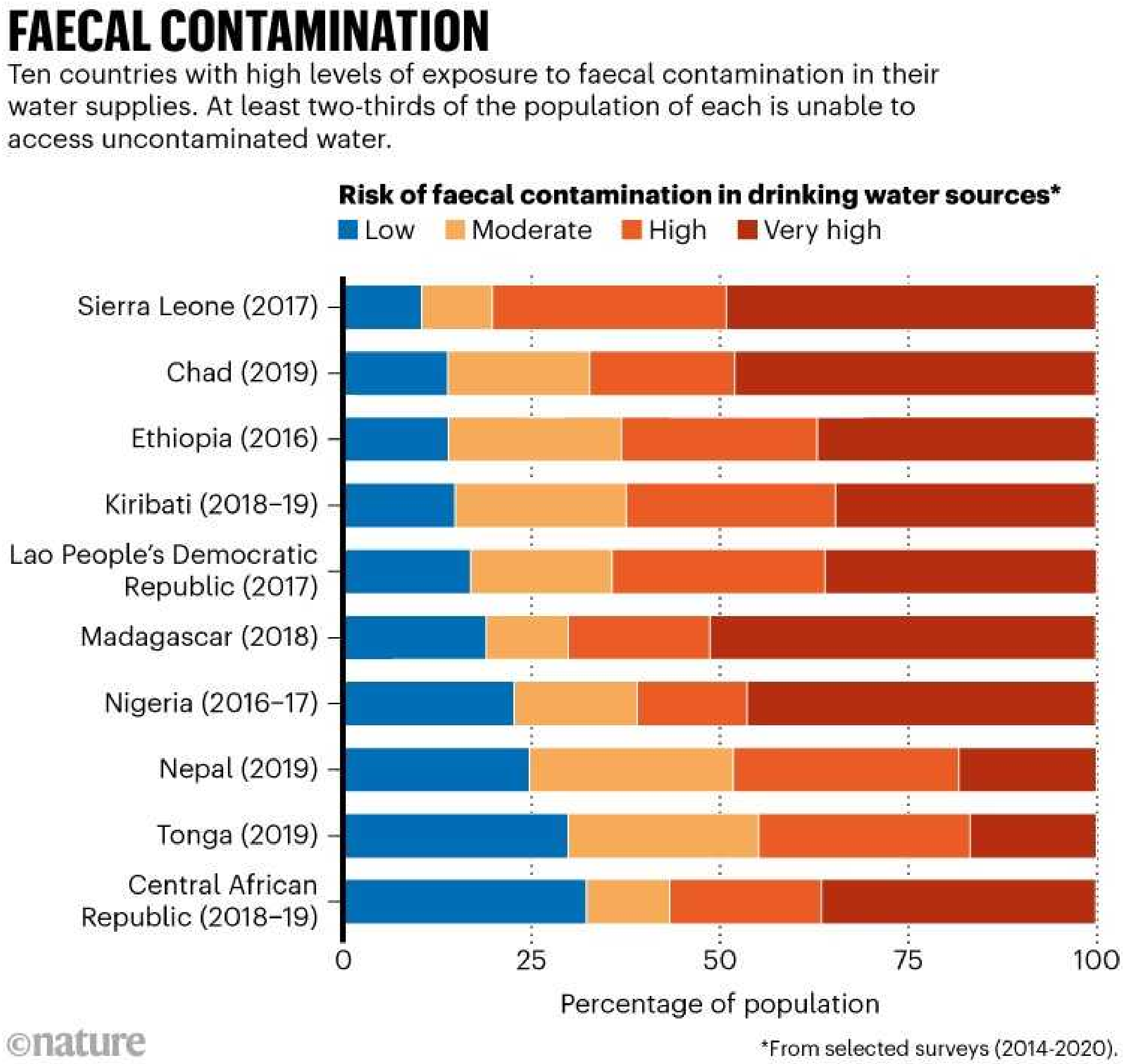
Will the conference lead to a legally binding water treaty?

The conference will produce a ‘water action agenda’. But this will involve “voluntary commitments”, says McDonnell. “There’s nothing binding, there is no equivalent to the Paris [climate] agreement.” Moreover, there’s no leading UN body that is responsible for implementing and monitoring progress for all water-related SDGs. “The conference finds itself in an institutional void,” says Ovink. “While we are now very busy with the water conference, we’re not very busy with water.”

But setting up a new treaty or institutional UN body will take many years. Instead, delegates will call for water to be prioritized in existing treaties and in the UN system.

Some countries will be calling for more funds, especially in the form of grants for projects such as desalination of seawater or wastewater treatment. Much existing international support is loans, says Omar Salameh, a spokesman for Jordan’s water and irrigation ministry, based in Amman. “However, loans exacerbate the financial pressures on already-struggling economies,’ he says.

UN secretary-general António Guterres is also expected to ramp up fundraising for his plan, announced at COP27, to create climate early-warning systems in all UN member states, so that countries are better prepared for extreme events. “Only half of our 193 members have proper early-warning services in place,” says Petteri Taalas, secretary-general of the World Meteorological Organization, based in Geneva, Switzerland, which is working with Guterres to implement the plan. “We need some US$3 billion during the coming five years,” adds Taalas. So far, around 10% of this has been raised, through different sources.



Source: WHO/UNICEF

Nature **615**, 774-775 (2023)

doi: https://doi.org/10.1038/d41586-023-00842-3

Updates & Corrections

- **Correction 22 March 2023**: An earlier version of the introduction to the graph 'The neglect of Africa' gave the wrong year for the comparison. This has now been corrected.

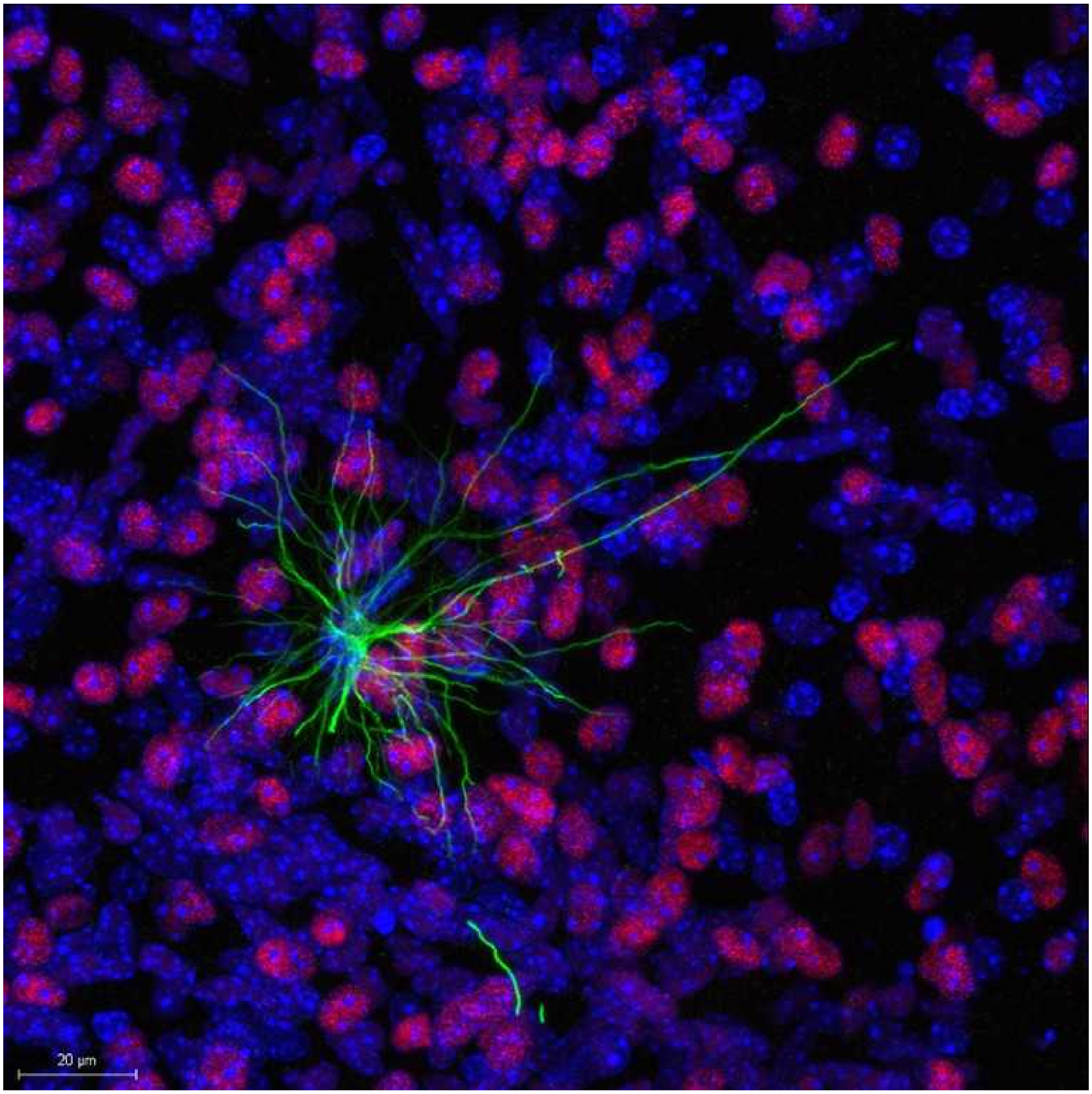
This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00842-3

- NEWS
- 16 March 2023

How stem cells make a human brain

Technical ‘tour de force’ allows researchers to trace the family tree of crucial brain cells.

- Heidi Ledford



A mouse brain (red- and blue-stained proteins) hosts a human astrocyte (green) grown from neural stem cells.Credit: Liu *et al./Cell (2023)*

In a technical "tour de force", researchers have analysed multiple traits of individual cells to pinpoint those that give rise to crucial components of the human brain.

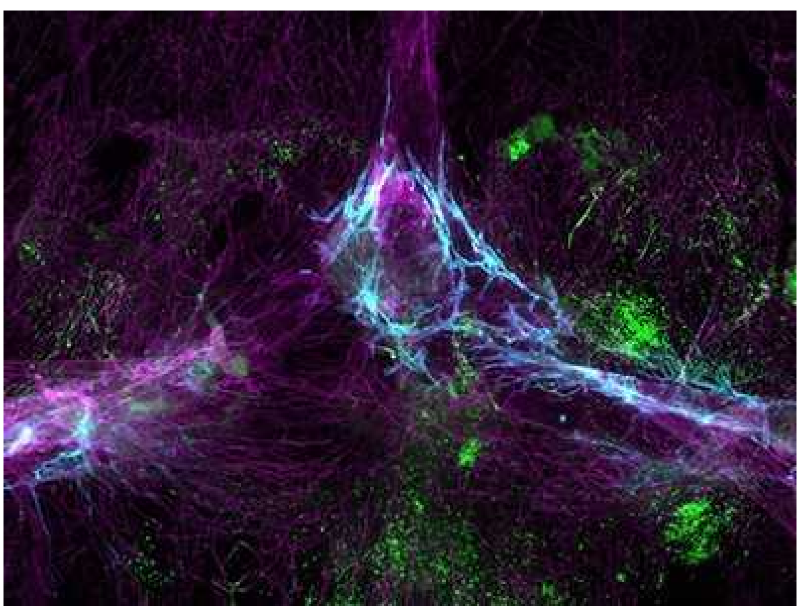
The analysis, published on 16 March in *Cell*[1], uses a combination of protein and RNA analysis to painstakingly purify and classify individual stem cells and their close relatives isolated from human brains. Researchers then

injected different types of cell into mice and monitored the cells as they divided and their progeny took on specialized roles in the brain.

The hope is that this study, and others like it, will illuminate how such developmental programmes go awry in neurological diseases — and how they can be harnessed to create new therapies. “The census of stem and progenitor cells in the developing human brain is really just beginning,” says Arnold Kriegstein, a developmental neuroscientist at the University of California, San Francisco, who was not involved in the research. “This work offers a nice window into some of that complexity.”

Cellular ensemble

The brain is an intricate symphony of different cells, each of which performs essential functions. Star-shaped cells known as astrocytes, for example, are important for supporting metabolism in neurons, and loss of astrocyte function is linked to neurodegenerative conditions such as Alzheimer’s disease. Oligodendrocytes are cells that create a protective, insulating sheath around the connections between neurons. When they are damaged — as in diseases such as multiple sclerosis — communication between neurons slows or stops altogether.



Guardians of the brain: how a special immune system protects our grey matter

To understand how such cells arise, stem cell biologists Irving Weissman and Daniel Liu and their colleagues at Stanford University in California harnessed new technology that would allow them to study the developmental destiny of individual cells taken from human brains.

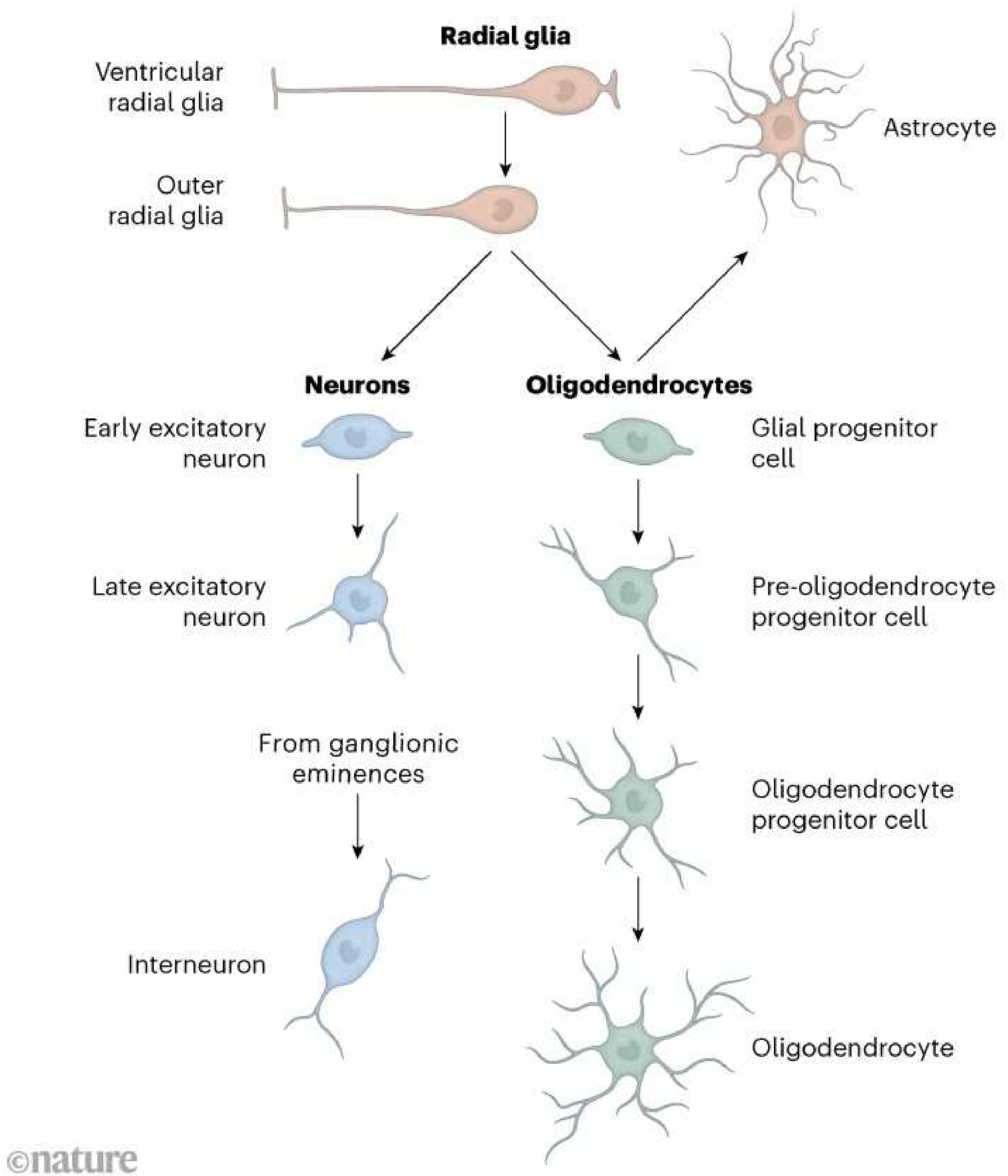
The team isolated brain cells from human fetuses that were 17–19 weeks old and tested the cells for a battery of 11 proteins on the cell surface, including six that are associated with neural cell development. They also analysed RNA levels as a measure of gene activity and used this information to purify 10 kinds of cell that are likely to give rise to astrocytes, oligodendrocytes and neurons.

The researchers then injected the purified cells into mouse brains. Six months later, they analysed those mice to find out where the cells and their descendants had migrated, and what cellular identities they had taken.

The approach enabled the team to define a new kind of progenitor cell that gives rise to glial cells, a grouping that includes astrocytes and oligodendrocytes (see 'How to make a brain'). These progenitors are derived from cells that are more sparse in mouse brains than in human brains, says Liu. "We think that this cell type might be important for specific adaptations that primate brains have made," he says.

HOW TO MAKE A BRAIN

Scientists have identified a host of stem cells and closely related cells that develop into three main types of brain cells: neurons, which carry electric impulses; astrocytes, which provide support to neurons; and oligodendrocytes, which supply the insulation on neurons. Cells of the same colour below have some of the same chemical markers on their surfaces.



Source: Ref 1.

The team also found that high levels of a protein called Thy1 are associated with cells in the oligodendrocyte lineage. This runs counter to previous findings, which suggested that Thy1 was a marker for neurons rather than oligodendrocytes, says Steven Goldman, a neurologist at the University of Rochester Medical Center in New York and head of central nervous system therapy at Sana Biotechnology in Cambridge, Massachusetts.

Such differences could also be the result of the new approach's improved resolution of different cell types, Goldman says, adding that the work is "a technical tour de force. … They upped the level of the field." Weissman says that the technique might be useful for studying other kinds of stem cell as well.

Complex brew for a brain

The study is an important contribution to the growing knowledge of the cellular lineages that give rise to the human brain, agrees Kriegstein. But he notes that the development of human stem cells in mice might not fully reflect how the cells would develop in human brains.

Goldman worries that the lineages derived by the analysis do not reflect the plasticity of neural cell development. Other research has found that some cells in the brain can begin down one lineage, only to change paths and emerge as an unrelated neural cell, he says. Liu and Weissman think that some of that apparent plasticity was instead an artefact of studying mixtures of cells, and might not be as pronounced when using the stringent selection criteria for purifying cells that they developed.

But Goldman suspects that other factors influence how committed neural cells remain to their lineage. "The nervous system is more complicated in terms of diversification," says Goldman. "There's still a lot to learn."

Nature **615**, 776-777 (2023)

doi: https://doi.org/10.1038/d41586-023-00762-2

References

1. Liu, D. D. *et al.* *Cell* https://doi.org/10.1016/j.cell.2023.02.017 (2023).

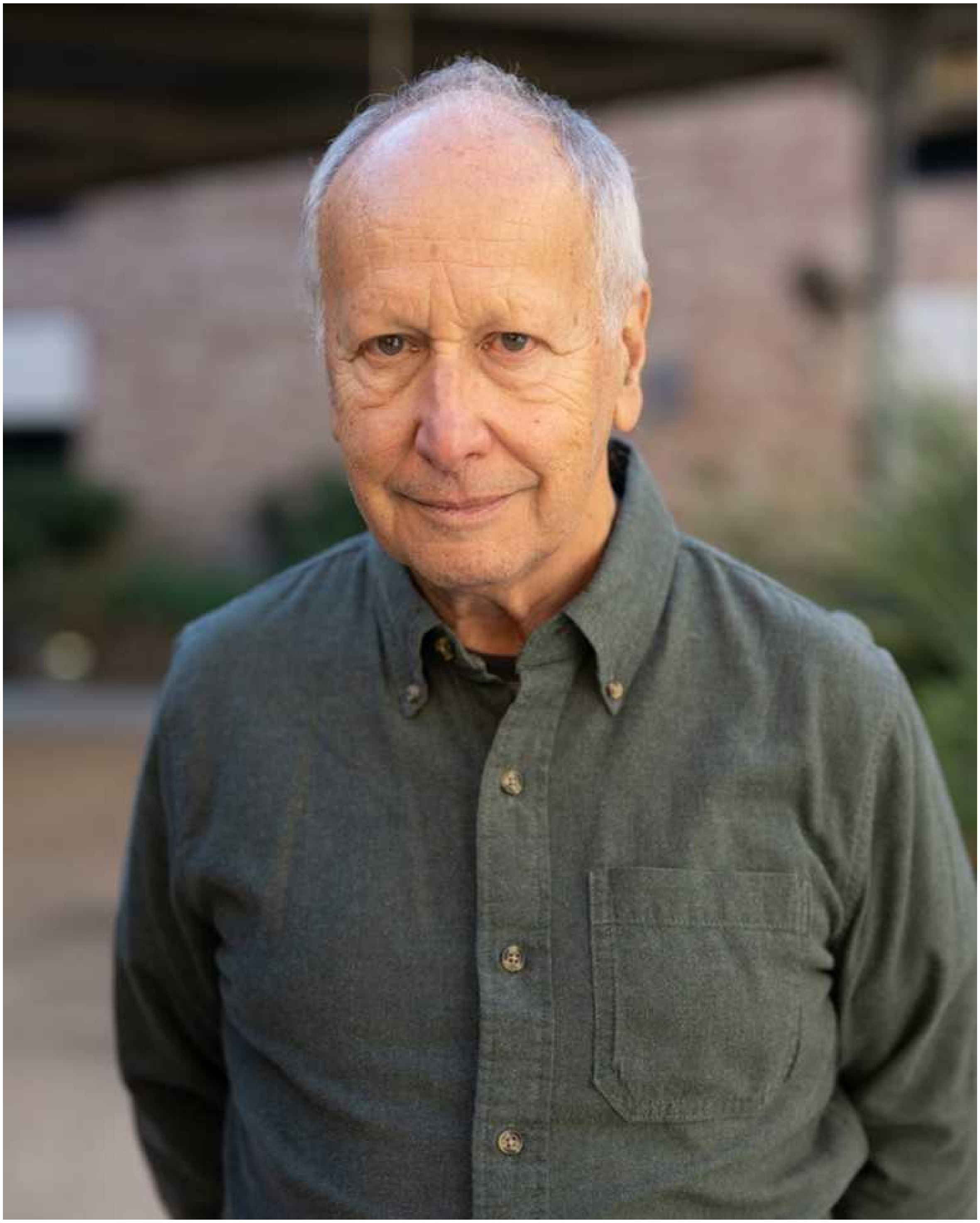
This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00762-2

- NEWS
- 22 March 2023

Abel Prize: pioneer of ‘smooth’ physics wins top maths award

Luis Caffarelli’s work includes equations underpinning physical phenomena, such as melting ice and flowing liquids.

- Davide Castelvecchi



Luis Caffarelli is the first person from South America to win the Abel Prize.Credit: Nolan Zunk/University of Texas at Austin.

Argentina-born mathematician Luis Caffarelli has won the 2023 Abel Prize — one of the most coveted awards in mathematics — for his work on

equations that are important for describing physical phenomena, such as how ice melts and fluids flow. He is the first person born in South America to win the award.

Caffarelli's results "are technically virtuous, covering many different areas of mathematics and its applications", says a statement by Helge Holden, a mathematician at the Norwegian University of Science and Technology in Trondheim who chairs the Abel Committee.

The winner says receiving the news was an emotional moment, because "it shows that people have some appreciation for me and for my science".

Caffarelli was born in Buenos Aires in 1948. After earning his PhD, he moved to the United States in 1973 and spent the rest of his career at various institutions there, including the Institute for Advanced Study in Princeton, New Jersey. He is now at the University of Texas at Austin.

Smooth solutions

Many of Caffarelli's most celebrated results have to do with the regularity — or lack thereof — of the solutions of so-called partial differential equations. These are among the most important equations in maths and physics. They involve multiple variables, typically time and one or more dimensions of space. Regularity means smoothness, so that the solutions don't form kinks or run away to infinity. In mathematical physics, solutions are often expected to be regular because the equations model phenomena at macroscopic scales — where things such as temperature or pressure distributions seem to be smooth and where infinities would be unphysical.



Virtuoso mathematician who re-shaped topology wins Abel Prize

In his first big breakthrough, Caffarelli generated a proof for the regularity of melting ice[1] — showing that the surface of a piece of ice remains mostly smooth as it turns into water. His result was refined in 2021 by Fields medallist Alessio Figalli, at the Swiss Federal Institute of Technology in Zurich, and his collaborators[2].

For another celebrated paper[3], Caffarelli joined forces in 1982 with the late Louis Nirenberg, the 2015 Abel laureate, and Robert Kohn, who is at New York University, to attack one of the most notorious problems in mathematical physics — the regularity of the motion of fluids, as described by the standard theory of fluid dynamics, known as the Navier–Stokes equations.

The trio proved a partial result, which is still essentially the best known solution to this day — and remains Caffarelli's proudest achievement, he says. But the full regularity is still an open question, and in 2000 it was included as one of the seven Millennium Problems by the Clay Mathematics Institute — which offers a prize of US$1 million to the first person to solve it. In principle, it is still possible that the equations fail to have regularity and they could instead allow the velocity of turbulent fluids to 'blow up' to infinity — making the theory a poor model for the real world, as Caffarelli explained in the official lecture introducing the Millennium Problem. In

more-recent years, Caffarelli has shifted to studying Monge–Ampère equations, which are important in geometry.

‘Fantastic intuition’

In a 2002 interview, Nirenberg said that Caffarelli had a “fantastic intuition”, which made it hard for collaborators to keep up with him. “He somehow immediately sees things that other people don’t see, but he has trouble explaining them,” Nirenberg said. Nirenberg also begged Caffarelli to publish more, because Caffarelli had a habit of putting into writing only a fraction of the ideas he would discuss with his peers.

“People say, ‘How did he come up with that one?’ He can come up with results that people didn’t anticipate,” says Holden.

Caffarelli describes his style of doing maths as “joyful”. “I enjoy collaborations, creativity, long-lasting friendship,” he says.



Soap-bubble pioneer is first woman to win prestigious maths prize

The fact that Caffarelli was raised and educated in Argentina gives the prize special significance, Holden adds. “I think this will have a great impact in the region,” he says. “It shows that talent comes from everywhere.”

Caffarelli will receive a medal and 7.5 million Norwegian kroner ($700,000) on 23 May at a ceremony in Oslo. Although it is modelled in part after the Nobel Prizes, which cite specific discoveries, the Abel Prize is seen by many as a 'career award'. Of the 26 laureates so far, only one, Karen Keskulla Uhlenbeck, is a woman.

Nature **615**, 777 (2023)

doi: https://doi.org/10.1038/d41586-023-00833-4

References

1. Caffarelli, L. A. *Acta Math.* **139**, 155–184 (1977).

2. Figalli, A., Ros-Oton, X & Serra, J. Preprint at https://arxiv.org/abs/2103.13379 (2021).

3. Caffarelli, L., Kohn, R. & Nirenberg, L. *Comm. Pure Appl. Math.* **35**, 771–831 (1982).

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00833-4

- NEWS Q&A
- 21 March 2023

‘Open for business’: risk-taking US health agency ready to spend $2.5-billion budget

One year after the launch of ARPA-H, *Nature* talks to director Renee Wegrzyn about her vision for the agency.

- Max Kozlov



Renee Wegrzyn is the first director of ARPA-H.Credit: Josh Reynolds/AP/Shutterstock

Launched one year ago by the administration of US President Joe Biden, the Advanced Research Projects Agency for Health (ARPA-H) has announced its first call for research proposals. The agency, armed with a US$2.5 billion budget, aims to shake up the conventional model of funding biomedical research — which some deem too slow and conservative — by funding high-risk, high-reward projects.

ARPA-H is meant to follow in the footsteps of the US Defense Advanced Research Projects Agency (DARPA), whose work laid the foundation for the modern Internet and GPS. But critics have questioned whether such a risk-based model will work for the life sciences, especially after administration officials announced that the agency would be housed in the US National Institutes of Health (NIH), a more conventional funder of biomedical research.

All eyes have been on Renee Wegrzyn, a biologist and former DARPA programme manager who Biden selected as ARPA-H's inaugural director, as she seeks to address these concerns and set the agency's agenda. She announced last week that the agency would have offices in three locations across the United States — one in the Washington DC area, and two others to be announced later this year.

Wegrzyn spoke to *Nature* about what's been going on behind the scenes and her vision for the future.

With ARPA-H turning one, how are you feeling about it? What progress has it made?

It's a super exciting time for us. I am feeling really great about having what we like to think of as our business team in place. Now we're bringing our technical team members [such as programme managers] on so that they can really hit the ground running. We finally have the critical mass that's allowed us to be open for business.

I saw that ARPA-H will have four focus areas. Tell me about them.

The first, health-science futures, is about tools, technologies and platforms. [People often ask me] what diseases ARPA-H focuses on, and I remind them that we're disease agnostic. We may advance the objectives of an initiative such as the Cancer Moonshot (Biden's $1.8-billion cancer-research programme), but it will be done in the context of a tool or platform.



[Billion-dollar US health agency gets new chief — but its direction remains in limbo]

Scalable solutions is making sure that we can scale some of the therapies and diagnostics that we intend to pursue, and it is making sure that we reach the American people at scale. Many rural communities are now facing hospital closings, and so they're farther from a centre of care. How do we reach the scale of getting technology and capability into people's homes?

Proactive health is about keeping people from becoming patients in the first place. Think detection and diagnostics, but also behavioural and social-science innovations that help people doing things such as quitting smoking or avoiding things that we know lead to disease over the long term.

The final area, resilient systems, is thinking at a much broader systems level. What are the things that we can integrate to help us be resilient against the next pandemic, economic crisis or climate crisis. Those are the investments that we just haven't seen at scale yet in other parts of government.

You’ve put out the first call for project proposals. What advice do you have for those applying?

This isn’t just a resubmission of a grant from somewhere else that you didn’t get funded. These are milestone-based projects that end with a specific objective, whether that’s a device, deliverables or some key insight, if it’s a very fundamental project. But we really want to see boldness. How will your work actually affect people’s health outcomes in the long run? Also understand that ARPA-H wants to be part of the conversation — it’s not simply ‘submit your idea and we fund it as is’. There’s often a negotiation and discussion of what we want to move forward with.

Why do you think the ARPA model is a good fit for health research, which, historically, has been approached more conservatively?



The rise of ‘ARPA-everything’ and what it means for science

mRNA vaccines had initially been developed with research at and funding from the NIH. But DARPA was really able to drastically accelerate getting those to market and demonstrating their safety in humans [by funding

vaccine makers Moderna and Pfizer]. It's a really great example of how an ARPA might be a catalyst. It's not meant to be the forever funder of these technologies, but that type of resourcing can really just quickly get a health advance to the public and into the hands of the people who need it most. I want to make clear, too, that under no circumstance does moving fast mean compromising on safety or efficacy — that is of utmost importance.

We're not going to be a grant-based organization. These are cooperative agreements, contracts and other transactional authorities, so that ARPA-H has a seat at the table.

Are there features of DARPA that you're hoping to incorporate into ARPA-H?

We're not just a copy–paste of the DARPA model. [Like DARPA,] we're not going to have a conventional peer-review process. The programme manager is going to be the decision-maker. (Programme managers at ARPA-based agencies help to set priorities and see projects through from idea to implementation.) We've also adopted the Heilmeier questions. George Heilmeier was an early director of DARPA, and he laid out eight questions that programme managers use to evaluate projects. We have adopted those, but we've adapted them to make them unique to us: we've added two more questions. One is focused on not only cost and accessibility, but also user experience — we want to make sure that we're not creating the next device or the next diagnostic that sits on the shelf. The other question we've added is: how might your new technology be misinterpreted or misunderstood?

When you were first appointed, I spoke to some researchers who were curious to see how the agency would incorporate equity. What are your plans?

When a programme manager comes into ARPA-H, we already know a bit about their idea from their job talk. They come, and they refine that idea into a full programme with milestones and metrics that we're going to pursue and

announce to the public. The last step before they announce it to the world is to pitch it to me, in the director's office, and to ask for a bank account. They have to really answer the 10 [fundamental] questions, including, how are you going to address accessibility and cost, and are you helping the patients that stand to benefit the most from your programme? If they can't answer those questions well, and they don't have a strategy and a plan for those issues, they don't get the bank account.

Nature **615**, 778 (2023)

doi: https://doi.org/10.1038/d41586-023-00817-4

This interview has been edited for length and clarity.

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00817-4

- NEWS FEATURE
- 17 March 2023
- Update 20 March 2023

Strife at *eLife*: inside a journal's quest to upend science publishing

Editors threaten to resign over 'no-reject' model that others see as the future of research journals.

- Alison Abbott

Last October, the pioneering life-sciences journal *eLife* introduced bold changes to its editorial practice — which some researchers applauded as reimagining the purpose of a scientific journal. From 31 January this year, *eLife* said, it would publish every paper it sent out for peer review: authors would never again receive a rejection after a negative review. Instead, reviewers' reports would be published alongside the paper, together with a short editorial assessment of the work's significance and rigour. Authors could then decide whether to revise their paper to address any comments.

The change followed an earlier decision by *eLife* to require that all submissions be posted as preprints online. The cumulative effect was to turn *eLife* into a producer of public reviews and assessments about online research. It was "relinquishing the traditional journal role of gatekeeper", editor-in-chief Michael Eisen explained in a press release, and "promoting the evaluation of scientists based on what, rather than where, they publish".



Michael Eisen, *eLife*'s editor-in-chief.Credit: HHMI

The transformation sparked enthusiastic praise — and sharp criticism. Some scientists saw it as a long-overdue move to empower authors. Others, including some of *eLife*'s academic editors (who are mostly senior researchers), weren't so happy. They worried it would diminish the prestige of a brand they'd worked hard to build, and some wrote privately to Eisen (in letters seen by *Nature*) to say they would resign if the plan was fully implemented. Amid the pushback, the journal postponed switching fully to its new process.

But the dispute only heightened. On 9 March, 29 *eLife* editors — including the journal's former editor-in-chief, Randy Schekman — wrote to Damian Pattinson, executive director of the journal's non-profit publisher, eLife Sciences Publications in Cambridge, UK, asking that Eisen be replaced "immediately". They added that they had no confidence in Eisen's leadership, because he had dismissed their concerns and had not considered compromise positions. One of the journal's five deputy editors had already stepped down from that leadership position, and "significant numbers" of reviewers and senior editors were "standing ready to resign", they wrote.

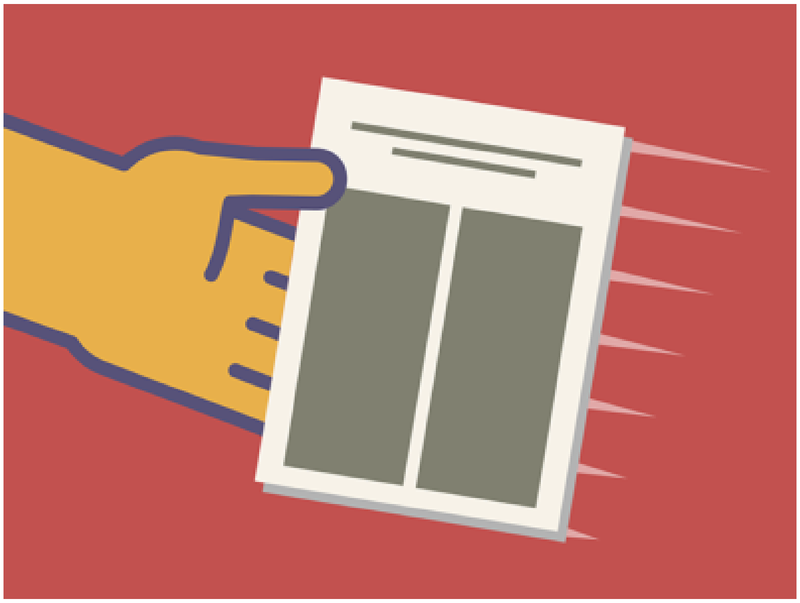
Eisen, a Howard Hughes Medical Institute (HHMI) investigator who works at the University of California, Berkeley, fired back publicly online, tweeting on 12 March that academics were “lobbying hard to get me fired”. He later deleted the tweet, but told *Nature* in an interview that “opposition to *eLife*’s model is driven fundamentally by powerful scientists not wanting to change a system that has benefited them and which they have sculpted to continue to reward them”. In response, Schekman and other authors stated that Eisen’s comments were “not true and do not reflect our legitimate concerns with the new model at *eLife*”.

Eisen says he thinks the dissent is small in scale. He and Pattinson say they did not dismiss concerns, but consulted on changes over two years with editors. “We see big swathes of enthusiasm among the community,” Pattinson adds.

The row highlights disagreements among researchers about the function of journals and peer review — and, potentially, about the future of science publishing. Some *eLife* editors argue that journals should use review to guide filtering and rejection of papers. But supporters of *eLife*’s changes see benefit in preventing peer review from serving as a prestige-gathering function, in which, by rejecting most of the manuscripts submitted to them, selective journals become perceived as arbiters of what work matters. “We rely too much on journal titles in judging people’s work,” Eisen says. “If we want to fix a bad system, we do have to break some eggs.”

What is a journal’s purpose?

When *eLife* was launched in 2012 with the financial backing of three powerful science funders — the Maryland-based HHMI, the UK Wellcome Trust and Germany’s Max Planck Society — it had the aim of being a non-commercial and academic-edited journal that would rival prestigious titles such as *Cell*, *Nature* and *Science*. (*Nature*’s news team is editorially independent of the journal’s research team.) Besides being open access, another of its key innovations was a collaborative system of peer review, where referees and a handling editor discuss comments together. The journal attracted dozens of working scientists as editors who triage submissions, with hundreds more scientists as reviewing editors.



[Open-access journal eLife announces ‘preprint first’ publishing model]

eLife had its eye on bigger changes, however. In 2021, the journal decided to publish only papers that were already preprints. This meant that delays in reviewing wouldn’t hold up an author from sharing their work. But even before Eisen and Pattinson joined, the journal had run a trial with more than 300 manuscripts to test the idea of ditching rejection after review. Its aim was to simply publish papers with reviews, author responses and editorial ratings. “The peer-review process does not need to end with a binary outcome of acceptance or rejection,” the journal wrote in a 2019 analysis of that work.

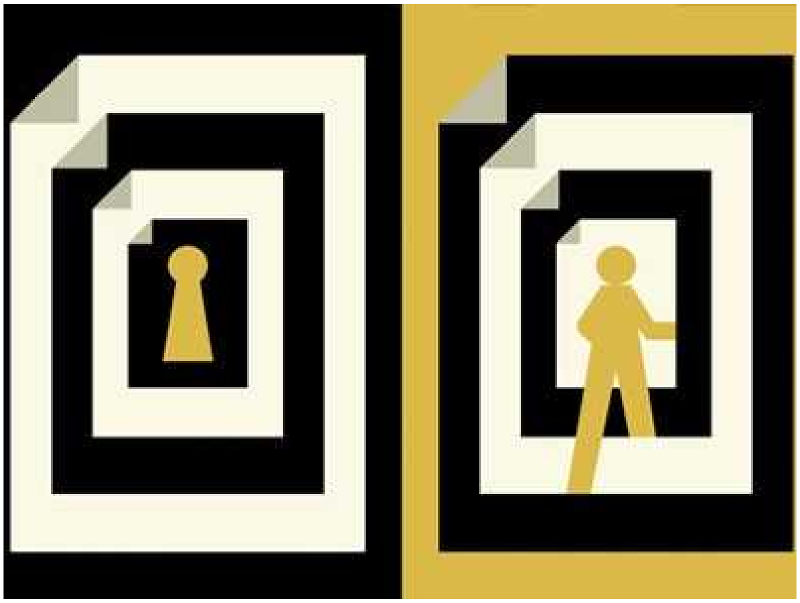
It was this idea that *eLife* instituted for all papers last October, with the addition that editors would also append a short summary assessment of the paper — giving readers a quick idea of its quality and significance. “This puts power back in the hands of the authors, who can then publish what they have, instead of having to do ever more experiments to satisfy reviewers,” says Eisen. The journal plans to charge US$2,000 for the process of arranging review on submissions; previously, its open-access publication fee was $3,000.

Some *eLife* editors are fully on board with the new system. “It’s the future, where science is going,” says senior *eLife* editor Panayiota Poirazi, a neuroscientist at the Institute of Molecular Biology and Biotechnology in

Heraklion, Crete. Among the journal's funders, HHMI says it fully supports the new policy. Wellcome says that it supports *eLife*'s publishing process, and the Max Planck Society told *Nature* it was still discussing the issue.

But other researchers have been openly critical from the start. In November, 47 editors wrote privately to Eisen asking for a rethink or for more time to experiment — perhaps running the new system alongside the conventional one, or creating a second journal in which to publish papers of less significance. They worried about harm to the journal's collaborative open-reviewing process, and that the quality of papers on the *eLife* platform would drop. With no possibility of rejection, some authors might choose to ignore reviewer comments or only superficially address them, they wrote — and that knowledge might discourage reviewers from producing detailed critiques. Responding to these concerns, Eisen and Pattinson say that they haven't seen such problems so far, although the project is in its early days, and that running two systems would reduce the chances of the new model's success.

Editors also argued that removing rejection-after-review meant more pressure on the gatekeeping step that remains in *eLife*'s system — the triage point where editors choose whether to send out a paper for review. That step had been "opaque and subject to errors in judgment", their letter stated, an issue that would become more consequential if later negative reviews could no longer lead to rejection. Editors might react by becoming more conservative and decide not to take a chance on manuscripts from less-well-known authors. But Eisen says that, in the new system, sending a preprint for review shouldn't communicate anything about its quality or importance: the reviews and editorial assessments do that instead. The guidance that editors should follow when deciding what to send for review is "can you generate high-quality and broadly useful public reviews of this paper?", he says.



[Will the pandemic permanently alter scientific publishing?]

In some countries, hiring and promotion decisions still rely heavily on journal titles in candidates' publication lists — something that is unlikely to change quickly, the editors added in their letter. They worried that scientists there would stop sending their manuscripts to *eLife*. Eisen, however, says that problematic reliance on journal titles will continue until there is an alternative system, such as *eLife*'s.

In a further private letter sent to Eisen in January, 30 editors said they would resign once the new policy was fully implemented.

The full scale of the discontent is unclear. Although Eisen and Pattinson say they've had broad support, Axel Brunger, a structural biologist at Stanford University in California, who initiated the first letter, says he reached out only to his colleagues in structural biology and neuroscience, and that nearly all agreed to sign up. "The concerns are widespread," he says.

One researcher who signed all three letters is neuroscientist Gary Westbrook at the Vollum Institute at Oregon Health & Science University in Portland. He is a vocal critic of what he sees as the monopoly that commercial journals have in science publishing, and says he signed "because I didn't think the new policy was realistic". Far from helping *eLife* as a non-profit,

high-quality alternative, he says, he thinks the model will diminish its impact.

Reviewing preprints

The concept of reviewing preprints is catching on in the life sciences. At least two dozen preprint-refereeing initiatives of various sizes have been launched in the past few years. The largest (apart from *eLife* itself) is Review Commons, launched in December 2019 by the California-based non-profit organization ASAPbio and EMBO Press. The latter runs five journals and is part of the European Molecular Biology Organization in Heidelberg, Germany. As a review-sharing collaboration between 17 journals from 6 publishers, including *eLife*, Review Commons uses EMBO Press editors to select referees for submissions. Authors can ask Review Commons to post reviews and any further author responses on a preprint server, or they can submit their paper, with reviews and responses, to any journal. More than 2,000 reviews of 540 articles have been run through this system.

The idea of ‘journal agnostic’ reviewing is still at proof-of-principle stage, says Bernd Pulverer, EMBO’s head of scientific publications. But he sees merit in having both peer-reviewed preprints and conventional journals, which, he says, provide “real added value in condensing and stratifying information”.

That view is shared by Maria Leptin, president of the European Research Council. “If I want to learn about a new field that is not core to my own, then I want a trustworthy source that filters for general interest,” she says. “*eLife* now does its filtering upstream, in a non-transparent, unaccountable way.”

The triage stage shouldn’t be seen as this kind of filter, says Eisen. “People are used to operating in a world where appearance in a journal tells you about the quality, audience or import of a study. This is precisely what we are trying to change,” he says. He argues that the short editorial summary *eLife* appends to its articles serve as quality guides for readers. They grade the significance of the findings (landmark, fundamental, important, valuable,

useful) and assess the strength of their support (exceptional, compelling, convincing, solid, incomplete, inadequate).

More consultation?

Endocrinologist Mone Zaidi at Icahn School of Medicine at Mount Sinai, New York City, is one of *eLife*'s four remaining deputy editors and has been trying to mediate the issue. He admires Eisen's vision, he says, "but any new, transformative change has to be done in a careful manner, with buy-in from the community".

Together with some of his colleagues, he is trying to persuade Eisen to slow down, to avoid mass resignations and to establish milestones to assess the effects the changes would have on the lives of working scientists. "There has to be consultation and risk-mitigation plans," he says.

The deputy editor who stood down, cell biologist Anna Akhmanova at the University of Utrecht in the Netherlands, shares Zaidi's view. She says she helped to develop the new system, but stepped down as deputy editor because it was being pushed through too fast. "We need evolution, not revolution — many small, careful steps to try to move the community towards what would be a better publishing system," she says.

Eisen says he has already responded to concerns by extending — for a short time — the deadline for the regular reviewing system. "We expect things to evolve in interesting ways as people start to see the advantages and opportunities of not making publishing decisions."

"*eLife* is doing a big and interesting experiment, however it works out," says stem-cell biologist Fiona Watt, a former *eLife* deputy editor who is now EMBO's director. "My sense as a scientist is that the publishing landscape is changing again."

Nature **615**, 780-781 (2023)

doi: https://doi.org/10.1038/d41586-023-00831-6

Updates & Corrections

- **Update 20 March 2023**: This article has been updated to highlight that *Nature*’s news team is editorially independent of the journal’s research team.

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00831-6

Chronic pain: the long road to discovery

By Lucy Odling-Smee

28 March 2023

Photography by Sam Murphy and Kevin Mohatt for *Nature*



The science is building and viable treatments are available.

But millions of people with chronic pain are still without help.

This article is also available as a pdf version.

Philip Kass spends 90% of his day lying on a twin bed in a sparsely decorated room that used to belong to his niece. He takes most meals with a

plate balanced on his chest, and he usually watches television because reading is too stressful.

"I'm barely living," he told me on a warm night in June last year.

Ever since a back injury 23 years ago, pain has been eating away at Kass's life. It has cost him his career, his relationships, his mobility and his independence.

Now 55, Kass lives with his sister and her family in San Francisco, California. He occasionally joins them for dinner, which means he'll eat while standing. And once a day he tries to walk four or five blocks around the neighbourhood. But he worries that any activity, walking too briskly or sitting upright for more than a few minutes, will trigger a fresh round of torment that can take days or weeks to subside.

"It's just paralysed me," he says.

Philip Kass has dealt with pain for more than two decades.

Philip Kass has dealt with pain for more than two decades.

Some of what Kass describes is familiar. I have been pinned to the floor by spinal pain several times in my life. In my twenties, I was immobilized for three months. In my thirties and forties, each episode of severe pain lasted more than a year. I spent at least another half decade standing or pacing through meetings, meals and movies — for fear that even a few minutes spent sitting would result in weeks of disabling pain. For years, I read anything I could find to better understand why my pain persisted.

The picture that emerged was complex and surprising. Over the past few decades, a growing body of evidence has indicated that the very machinery that processes pain can help to sustain the sensation or make it worse. Some researchers have explored unexpected interactions between the immune and nervous systems, showing, for instance, that inflammation, long considered a provocateur of pain, might also be crucial for resolving it. Others have shown how depression, anxiety and other kinds of emotional distress can both feed — and feed off — the experience of pain.

Although there are treatments that acknowledge this multifaceted, biopsychosocial nature of pain, medical practice hasn't kept pace with the science — not by a long shot.

Around the world, physicians are prescribing the same kinds of intervention for chronic pain conditions that they have offered for decades. Many people aren't getting treatment at all. And in North America and Australia, opioid overprescription and misuse have resulted in tight restrictions on one of the main tools long used by clinicians for managing pain. Kass, and the half-dozen other people I interviewed who live with pain, have spent years bouncing from doctor to doctor, trying to find someone who can eradicate their torment, or at least lessen it. And they're not alone.

In the last major survey, conducted in 2016, around 20% of adults in the United States — about 50 million people — had experienced pain on most days or every day in the previous 6 months. Around 8% — nearly 20 million people — had high-impact chronic pain, the kind that interferes with work or day-to-day activities[1]. Similar numbers have been reported from surveys conducted in Europe.

Although the pain research field is small, and fractured between different specialities, some researchers and clinicians argue that the knowledge and tools are already available to treat people with chronic pain conditions more efficiently and effectively than has been done in the past. What's needed, they say, is the will to get there — from both the medical establishment and society at large.

"We have a lot of treatments and approaches that already can make an impact, but we need to get them into people's hands, and we need to pay for it," says Sean Mackey, a physician-scientist at Stanford University in California.

The body's alarm

Pain often seems like a simple experience: grab a hot pan, and a harm alarm will sound. But in many cases where pain persists, the relationship between pain and harm is anything but simple.

In 1996, when Kass was 28, he was working as an acrobatics instructor with a travel and tourism company in the Bahamas. One morning, after a day of catching people on the flying trapeze, he was awoken by excruciating pain in his lower back.

He quit his job and flew back to the United States, where magnetic resonance imaging (MRI) revealed damage to a cartilaginous disc in his lower back, between his L5 and S1 vertebrae.

It's impossible to know exactly what was going on in Kass's body in the first hours and days after his injury. The specialists he saw in the United States told him that fragments of material from the damaged disc were probably pressing on his spinal nerves. But assuming Kass had experienced some kind of tissue damage, inflammation resulting from that would almost certainly have altered his pain-processing machinery.

Researchers have known for decades that tissue damage can alter the sensitivity of neurons in the peripheral nervous system (all of the nerves that exist outside the spinal cord, brainstem and brain), as well as how the central nervous system interprets signals.

After a sunburn, for instance, warm water that felt pleasant the day before might seem scalding. This happens because the inflammation caused by the sunburn alters the sensitivity of nerves in the peripheral nervous system known as nociceptors, which recognize noxious stimuli — a phenomenon called peripheral sensitization. Similarly, days after surgery, well away from the site of the incision where there is no inflammation, a light touch of the skin might hurt. That's thanks to changes in the central nervous system. Multiple mechanisms drive this process, known as central sensitization, but in this case, firing from sensory neurons activated by innocuous stimuli is now perceived as pain.

> "Evolution took a piece of string and randomly … knotted it into a hellacious ball of yarn."

In most people, most of the time, peripheral and central sensitization are temporary and adaptive. They prevent people from doing more harm to

damaged tissues. But what if that amplified sensitivity persists after wounds have healed, or even in the absence of any detectable tissue damage?

Various studies, mainly in animals, have identified dozens of pathways and cell types involved in peripheral and central sensitization. Following damage — the cutting or crushing of a rat's sciatic nerve, say — a menagerie of cells get activated around the nociceptors, and release factors that make the neurons more sensitive. These include macrophages, neutrophils, T cells and B cells, as well as glial cells, which are non-neuronal cells that support and protect neurons.

It's a remarkably complex set of interactions, says Jeffrey Mogil, a neuroscientist at McGill University in Montreal, Canada. "Evolution took a piece of string and randomly — because it had millions of years — knotted it into a hellacious ball of yarn," he says.

And the communication between nociceptors and immune cells goes both ways. In certain contexts, pain-triggering neurons can block or ramp up the activities of neutrophils and other types of immune cell[2]. "The nervous system doesn't even have to go via the brain; it just signals directly to the immune system in the periphery," says Isaac Chiu, a neuroimmunologist at Harvard Medical School in Boston, Massachusetts.

In people with skin conditions such as eczema, where there is ongoing inflammation, these reciprocal interactions between the immune system and nociceptors might help to drive persistent inflammation, and with that, persistent pain.

Even a person's microbiome, which interacts with both the immune and nervous systems, could play a part in certain pain conditions. Several groups, for example, are exploring the use of probiotics to treat people with irritable bowel syndrome (which causes abdominal pain)[3].

Another emerging idea is that some immune-system processes that drive sensitization might also be important in driving pain away. Last year, scientists at McGill University published an analysis of gene-expression patterns in people with lower-back pain. Although clinical trials are needed

to verify the results, their data indicate that if inflammation is blocked by drugs, neutrophils don't do what they are supposed to do to resolve pain[4].

This flies in the face of expectation, says Clifford Woolf, a neuroscientist also at Harvard Medical School — and the first to demonstrate central sensitization[5]. Physicians have long prescribed anti-inflammatory drugs on the premise that if pain is allowed to persist, it might become chronic. "This paper suggests the complete unexpected opposite, which is that the inflammation is actually helping," Woolf says.

Pain's amplifiers

Any one person's chronic pain develops in its own particular way.

After two years spent mainly lying on his back, Kass had surgeries to fuse several vertebrae. His condition stabilized enough for him to start a handyman business. He was still in pain most of the time, but as long as he limited his hours and the kinds of job he did, it was manageable.

Then, in 2008, for no apparent reason, Kass's pain worsened. To relieve it, he started taking Norco, a mix of hydrocodone and paracetamol (acetaminophen). He had another operation to insert a device that delivers pain medication directly to the spinal fluid. Over the years, he has had three steroid injections at different points along his spine.

Nothing worked. By 2015, Kass had abandoned his business. Exercise was no longer worth the risk. There'd be "this awful lag", he says, where he'd do something — ride his bicycle for a few blocks, for example — only to be incapacitated the following day.

It's anyone's guess as to whether immune-system processes, nerve damage or some other pathology in his lower back are helping to keep Kass in pain today. But neuroscientists have been uncovering all sorts of ways in which, over time, the experience of pain can be sustained and even amplified.

> "There's no doubt that the context in which you experience pain … will dramatically influence your perception of it."

In the past few decades, brain-imaging studies have revealed changes to different areas of the brain in people who have experienced prolonged pain — including in the limbic system, the part involved in behavioural and emotional responses[6]. And experiments using animal models suggest that some of the neural networks that get reorganized by chronic pain might, in turn, affect how pain is perceived. In 2019, researchers at Stanford University used a technique called chemogenetics to engineer mice so that their neuronal activity could be finely tuned using various drugs. When the researchers turned off certain neurons in a part of the limbic system called the amygdala, the mice could still sense pain, but behaved as if that pain was less bothersome[7].

These findings raise questions about whether a new type of painkiller could be developed that acts on the unpleasantness of pain rather than on pain sensation, explains Grégory Scherrer, who led the study and is now at the University of North Carolina at Chapel Hill.

Another area of work involves looking for associations between certain properties of people’s brains — as shown by functional MRI (fMRI) — and how they had previously responded to questionnaires assessing their levels of depression or anxiety. These studies have revealed that attention, expectations, anxiety, depression, catastrophic thinking and more can all influence pain perception[8].

Anyone who has been in severe pain knows that with enough discomfort, one thought will dominate all others: ‘make it stop’. But at least in certain situations, there might be some truth in the saying that attitude is everything. “There’s no doubt that the context in which you experience pain — your expectations and your attentional focus on it — will dramatically influence your perception of it, through real physiological neural networks,” says neuroscientist Irene Tracey, vice-chancellor of the University of Oxford, UK.

That context can include the people around you and the attitudes they bring. Among the things that can heighten anxiety and depression — which can, in turn, escalate pain — is scepticism from doctors, friends, partners and family. Many people with chronic pain face biases, which can be

exacerbated by factors such as race, stigma around opioid use and misuse and the source of pain not being visible or obvious.

Quána Madison, in severe pain since 2016, says her concerns about pain have sometimes been met with suspicion.

Quána Madison, in severe pain since 2016, says her concerns about pain have sometimes been met with suspicion.

Quána Madison, a 42-year-old artist living in Denver, Colorado, has felt severe pain in different parts of her body every day since 2016. In that one year, she had surgeries to remove her breasts, womb and ovaries. These were followed by breast reconstructions and another emergency operation to repair a potentially life-threatening complication of hysterectomy.

Madison, who is African American, has had to deal with the pain, but also with suspicion and bias from others. During one of her visits to an emergency department in 2017, a nurse, who had been trying to draw her blood, called the police after Madison warned her that because of her heightened sensitivity to pain, she might scream if the nurse didn't use a small-gauge needle and a hot pad.

Predictors of pain

Faced with unrelenting pain and other people's doubts, it's easy to start ruminating: 'Why me? What is it about my body, my brain, my past, that makes this pain soar, not subside?'

Researchers have begun to unpick some of the factors that correlate with a heightened risk of chronic pain using genomics, epidemiological surveys and brain imaging.

Studies of twins indicate that the heritability of common chronic pain conditions (irritable bowel syndrome, back and neck pain and migraines) ranges from around 25% to 50% (ref. [9]). But genomics studies, mainly involving people experiencing migraines, have generally uncovered gene variants that, individually, have small effects on a person's risk of developing chronic pain[10].

Various risk factors for chronic pain have been identified by epidemiological surveys, in which thousands of people are given questionnaires. The main factors that affect risk are: age; sex; socio-economic status; levels of anxiety, depression, sleep and physical activity; and body mass index[11]. For reasons that have yet to be worked out, persistent pain is more likely to be reported by people identifying as women than by those identifying as men. In the case of fibromyalgia, a disorder characterized by widespread musculoskeletal pain and fatigue, and problems with sleep, memory and mood, women are nine times more likely to have the condition than are men[12]. Also, people's likelihood of developing a chronic pain condition increases as they age, until about 60.

> "If you do make someone's pain dramatically better … you often make anxiety, depression, catastrophizing dramatically better."

Separating cause and effect is a challenge. Are people who report high levels of depression and anxiety more likely to develop a pain condition? Or is their pain making them depressed and anxious?

According to Daniel Clauw, a physician-scientist at the University of Michigan in Ann Arbor, many studies show that "if you do make someone's pain dramatically better", say with knee replacement surgery, "you often make anxiety, depression, catastrophizing dramatically better".

But there might also be something about people's brains that predisposes them to chronic pain.

Using fMRI, Tracey and others have assessed the brain properties of groups of people with a pain condition (back pain or osteoarthritis, say). After surgery to treat the condition, or after a certain amount of time has elapsed, only some people in each group continue to have pain. Across multiple studies, researchers have been able to make predictions about who will fall into which group, using the patterns of neural activity in people's brains when they first enrolled.

This January, Tracey and her colleagues assessed people who were about to receive chemotherapy for various types of cancer. Around 30% of cancer survivors develop a chronic pain condition called chemotherapy-induced

peripheral neuropathy. In this study, the researchers were able to use people's fMRI brain-imaging data, collected before chemotherapy, to predict which group they would fall into[13].

Tracey says that the hope is to identify which individuals are most at risk of neuropathy from chemotherapeutic agents, and then to adjust the treatments to reduce the risk.

Targeting treatments

The approaches currently used to treat common pain conditions generally entail medications and interventions such as surgical procedures or nerve blocks — the injection of a local anaesthetic close to a nerve or group of nerves. There are also behaviourally and psychologically oriented approaches, such as cognitive behavioural therapy, which involves trying to alter patterns of thinking and behaviour around the pain, and physical therapy to increase levels of activity and functionality.

"All these things help some people some," says David Clark, an anaesthesiologist at Stanford University School of Medicine.

Clinicians and others estimate that treatments available today can help one in three to one in four people, such that the levels of pain they report could decrease by about 30–50%.

Some researchers think that better efforts to match people with pain to the right treatments could make these options more effective.

One example comes from a multi-year project called the Multidisciplinary Approach to the Study of Chronic Pelvic Pain (MAPP) Research Network. Data from this study[14] suggest that people diagnosed with bladder pain syndrome (formally known as interstitial cystitis), but who have pain confined to their pelvis, essentially have a different disease from those given the same diagnosis but who have widespread pain throughout their body.

In neuroimaging, the brains of those with confined pelvic pain appear healthy, whereas people with widespread pain have brains that look similar to those of individuals diagnosed with fibromyalgia, explains Clauw, who

was involved in the study. This suggests that the group with widespread pain might respond better to drugs than to physical therapy for the pelvis[14].

"Within the next couple of years, we'll take a lot of commonly used treatments that work in one out of three individuals, and just by getting smarter about who we give those treatments to, they will work in one out of two people," says Clauw.

> "I do meditation. I use aromatherapy. I do abstract painting. I try to do anything possible to help me cope every day with my pain."

Similarly motivated by the idea of better matching treatments to individuals, Mackey at Stanford is using a US$12-million grant from the US National Institutes of Health to develop biomarkers for pain.

Painting is one of the ways that Madison copes with her pain.

Painting is one of the ways that Madison copes with her pain.

Around ten years ago, to improve care for people with chronic pain, Mackey created a digital platform called CHOIR (Collaborative Health Outcomes Information Registry). This characterizes people according to their physical, psychological and social functioning — largely on the basis of clinicians' reports and participants' responses to health questionnaires. Ultimately, Mackey's goal is to incorporate brain-based biomarkers of pain into the CHOIR system, as well as other information distilled from metabolomics, proteomics, genomics, the microbiome and even from wearable fitness trackers such as Fitbit.

Mackey and his colleagues are still trying to map associations between the vast array of biomarkers that can be measured and the experience of pain. But he's convinced of the promise of the basic approach: "I'm growing more and more optimistic that we're going to be able to do this," he says.

Most researchers and clinicians agree that better-tailored treatments could help. But many people with persistent pain have difficulty accessing any care at all.

"When I see the challenges facing patients," says Mackey, who has co-led a US effort to change how people in pain are assessed and treated, "I see this as more of a societal implementation issue."

Thanks to clinicians discovering that drugs developed for other conditions, such as seizures or depression, can help treat pain, there are actually many more treatments for chronic pain today than there were 20 years ago. Among the ones Mackey lists are more than 200 medications, most of which are non-opioids; mind-body therapies, such as acceptance and commitment therapy; and a host of procedures such as those involving spinal-cord stimulators — implanted devices that send low levels of electricity into the spinal cord.

They just aren't getting to the people who might be helped by them — at least to some degree.

Madison, and some of the other people I spoke to who live with pain, told me similar stories: unable to get much assistance from the health-care system, they have eventually cobbled together their own forms of treatment, born of years of trial and error.

"I do meditation. I use aromatherapy. I do abstract painting. I try to do anything possible to help me cope every day with my pain," says Madison.

Clinicians say a host of changes are needed to improve the situation. These include a shift in attitudes towards chronic pain conditions; insurance companies being persuaded to cover integrative care involving multidisciplinary teams; better education and training about pain in medical schools; and a lot more investment in the problem.

Researchers who continue to chip away at the mechanisms underlying chronic pain conditions could gradually change the narrative around chronic pain. And it's not unthinkable that, some day, providers of pain care will be trained in a very different way from how they are today. But right now, "it's really hard to find good quality care for people with chronic pain", acknowledges Clauw.

Breaking point

In the winter of 2020, whatever it was that had been holding Kass together through 23 years of pain broke down. In the mornings, he would cry. At some point, the crying turned into screaming. One of the tenants in Kass's building in Berkeley, California, closed his shop for a few weeks because of the noise.

That tenant, who was also Kass's friend, would visit him now and again to try to help. But one day in early 2021, Kass overdosed on medication with the intention of killing himself.

Kass thinks he was unconscious for about 24 hours before his friend found him.

A team of emergency doctors at Alta Bates Summit Medical Center in Berkeley revived him, and kept him under observation for three days.

If you or someone you know is contemplating suicide, please reach out. You can find help through a suicide-prevention line: see findahelpline.com for information.

"It was awful," he says. "I was scared. I was in terrible pain. And I couldn't leave the hospital."

After days of Kass begging the hospital staff to give him pain medication, a psychiatrist prescribed him an opioid called buprenorphine. Along with mirtazapine, other antidepressants and an antipsychotic called Latuda (lurasidone), he's still taking the opioid: eight milligrams three times a day.

Kass worries, however, that because of the kind of health insurance he has and doctors' wariness about prescribing opioids in the United States, his prescription could be stopped any day. "So there's that fear," he says. "Without the buprenorphine, I'm back to screaming every day."

Pain is a universal experience. Yet it's hard to comprehend what being ruled by fear and wracked by a continually sounding alarm can do to a person — even for those of us who have, at some point, experienced a long stretch of severe pain.

For almost all of 2020 and some of 2021, I limited my standing or walking to 5–10-minute sessions. After each attempt to move — often walking the block just a few streets from where Kass lives now — I would lie down for an hour. While on my feet, I wouldn't let my sons, then aged six and eight, touch me in case their pulling triggered another two-week-long flare. I wore ski gloves while lying on the floor because cold worsened the stabbing in my arms. And at night, I couldn't sleep for more than a two-hour stretch.

In that winter of 2020, I had screamed too.

Although I still have pain every day, and periodic flares, I've clawed a full life back little by little, using approaches similar to those a multidisciplinary UK team had introduced me to more than a decade ago. I've used medication, meditation and psychotherapy. I've used cognitive behavioural therapy to monitor my thinking around the pain, and eventually to alter it. I've used a timer to pace my activities, initially increasing the spells on my feet by a few seconds. I've used physical therapy and exercise to recondition my body — and to keep pushing back against the 'something is wrong' alarm that's been endlessly looping.

But pain takes a different course for every person. And as I listen to Kass, whose skin is so pale it's almost translucent, and who sometimes strains for a date or a word more than I'd expect for a man his age, I struggle to understand why the health-care systems of the world's richest countries are not yet using all the knowledge that science has provided to bring more help to him and millions of others living with pain.

Lucy Odling-Smee is a Senior Comment Editor for Nature in San Francisco, California.

References

1. Dahlhamer, J. *et al. MMWR Morb. Mortal. Wkly Rep.* **67**, 1001–1006 (2018).
2. Michoud, F. *et al. Nature Biotechnol.* **39**, 179–185 (2021).

3. Lagomarsino, V. N., Kostic, A. D. & Chiu, I. M. *Neurobiol. Pain* **9**, 100056 (2021).
4. Parisien, M. *et al. Sci. Transl. Med.* **14**, eabj9954 (2022).
5. Woolf, C. J. *Nature* **306**, 686–688 (1983).
6. Wood, J. N. (ed.) *The Oxford Handbook of the Neurobiology of Pain* (Oxford Univ. Press, 2018).
7. Corder, G. *et al. Science* **363**, 276–281 (2019).
8. Tracey, I., Woolf, C. J. & Andrews, N. A. *Neuron* **101**, 783–800 (2019).
9. Nielsen, C. S., Knudsen, G. P. & Steingrímsdóttir, Ó. A. *Clin. Genet.* **82**, 331–340 (2012).
10. Hautakangas, H. *et al. Nature Genet.* **54**, 152–160 (2022).
11. Macfarlane, G. J. *Pain* **157**, 2158–2159 (2016).
12. Yunus, M. B. *Curr. Rheumatol. Rep.* **3**, 128–134 (2001).
13. Seretny, M. *et al. Br. J. Anaesth.* **130**, 83–93 (2023).
14. Kutch, J. J. *et al. Pain* **158**, 1979–1991 (2017).

SPRINGER NATURE
Springer Nature

- Privacy Policy
- Use of cookies
- Legal notice
- Terms & Conditions
- Accessibility statement

© 2023 Springer Nature Limited. All rights reserved.
TopBuilt with **Shorthand**

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00869-6

Books & Arts

- **How Nancy Hopkins and her tape measure revealed the extent of sexism in science** [27 Mar 2023]
  • When the MIT molecular biologist quantified how much more lab space men were given, she unleashed a US-wide reckoning about how women are held back in academia.
- **When the immune system bites back, and biodiversity v. capitalism: Books in brief** [03 Feb 2023]
  • Andrew Robinson reviews five of the best science picks.

- BOOK REVIEW
- 27 March 2023

How Nancy Hopkins and her tape measure revealed the extent of sexism in science

When the MIT molecular biologist quantified how much more lab space men were given, she unleashed a US-wide reckoning about how women are held back in academia.

- Alexandra Witze



Nancy Hopkins showed how the Masschusetts Institute of Technology systematically awarded more lab space to men.Credit: Everett Collection Inc/Alamy

The Exceptions: Nancy Hopkins, MIT, and the Fight for Women in Science *Kate Zernike* Scribner (2023)

Nancy Hopkins's professional career has been partly defined by the 'great men of biology' she has worked with. Hopkins, a molecular biologist at the Massachusetts Institute of Technology (MIT), recounts being mentored by James Watson, yelled at by Eric Lander, slighted by David Baltimore and groped by Francis Crick. But her actual legacy rests in her scientific achievements in cancer biology and zebrafish genetics — and in the attention she drew to discrimination against women in science.

One night in 1993, seething at the fact that her male colleagues routinely took up more than their fair share of laboratory space, Hopkins got out a tape and measured the dimensions of each lab, office and equipment room in her building to quantify the differences between the spaces granted to women and men. Her work led to a seminal 1999 report from MIT in which the university admitted to having discriminated against female members of its science faculty for decades. That led to a nationwide reckoning about how institutions of higher education routinely held back women in science.

In *The Exceptions*, journalist Kate Zernike details Hopkins's journey from a young student convinced that academia was a meritocracy to a seasoned faculty member who saw that the opposite was true. Zernike's account details more than just the journey of one scientist — it provides a deeply researched dive into the history of gender discrimination in US higher education. The 'exceptions' of her title are the exceptional women who pushed through discrimination in science to have accomplished careers, as Hopkins did. Even though the main events transpired decades ago, they remain remarkably relevant today given the sexism, racism and other injustices that still permeate academia.



Revealed: the pay bump for being a straight, white man in US science

MIT was perhaps an unlikely place for this reckoning to unfold. Located in Cambridge, Massachusetts, a stone's throw from the elite bastion of Harvard University, MIT was supposed to be a meritocracy by contrast, a place where proficiency in engineering and science levelled the playing field of privilege. MIT's first female student graduated in 1873, a century before Harvard fully deigned to admit women.

But the few women who actually enrolled at MIT faced an uphill battle. Campus housing was nonexistent at first; a women's dormitory was added in the 1960s, but it was still insufficient. Women were sidelined, propositioned and assaulted as they tried to work and study. As late as 1985, three years before I entered MIT as an undergraduate, the university showed pornographic films in large auditoriums as a start-of-semester tradition.

After finishing her postdoctoral studies with Watson and Robert Pollack at the Cold Spring Harbor Laboratory in Laurel Hollow, New York, Hopkins arrived at MIT in 1973 as only the second woman ever on the faculty in the biology department. She was going against the advice of Nobel laureate geneticist Barbara McClintock, who warned her not to take any university job because of the likelihood of discrimination. She had already made important discoveries in cell and cancer biology, and was about to pioneer work on the role of retroviruses in cancer. Yet she was still frequently told

that by working in science she would be taking a job from a more deserving man, and that she could not teach genetics to undergraduates because they would not trust scientific information coming from a woman.

Hopkins's previously unpublished notes from this time reveal a searing picture of injustice. She is deliberately demoted on a priority list for tenure despite her formidable application; excluded from the profits of a biology course that she co-developed; and sexually assaulted by a colleague. After two decades at MIT, she finally realizes that science is not an unblemished search for the truth but rather a scrum of salesmanship full of competition, greed and harassment in which men have the upper hand. It all culminates, in Zernike's narrative, in Hopkins's infamous tape measurement.



To advance equality for women, use the evidence

This might have remained just one dramatic late-night episode, were it not for Hopkins joining forces with 15 other women on the MIT science faculty to bring the discrimination to light. They went to the science dean and, with his encouragement, produced the 1999 report on the status of women at MIT. The university released the study, which concluded that the discrimination had persisted across generations of female faculty members. Within days, Hopkins was giving media interviews around the world.

Zernike covered that story as a reporter at the *Boston Globe*. She now illuminates its backstory, placing the MIT experience in the history of US higher education, from how universities struggled to cope with the feminist revolution of the 1960s to the Title IX law of 1972 that banned discrimination on the basis of sex in federally funded institutions.

It is a single story acutely told, with a historical context that enriches and deepens its narrative. Zernike does not, however, address gender discrimination at other institutions or include context from academia outside the United States. Issues of intersectionality, in which gender, race and other factors combine to amplify discrimination, are explored, but not at length.

So why tell Hopkins's story now? The tape measure is now in the MIT museum, but that doesn't mean things have changed. A study released in January by the Scripps Institution of Oceanography in San Diego, California, found that its female scientists have, on average, just half as much research space allotted to them as their male counterparts. Thirty years after Hopkins's measurements, discrimination in academia remains alive and well.

Nature **615**, 787-788 (2023)

doi: https://doi.org/10.1038/d41586-023-00887-4

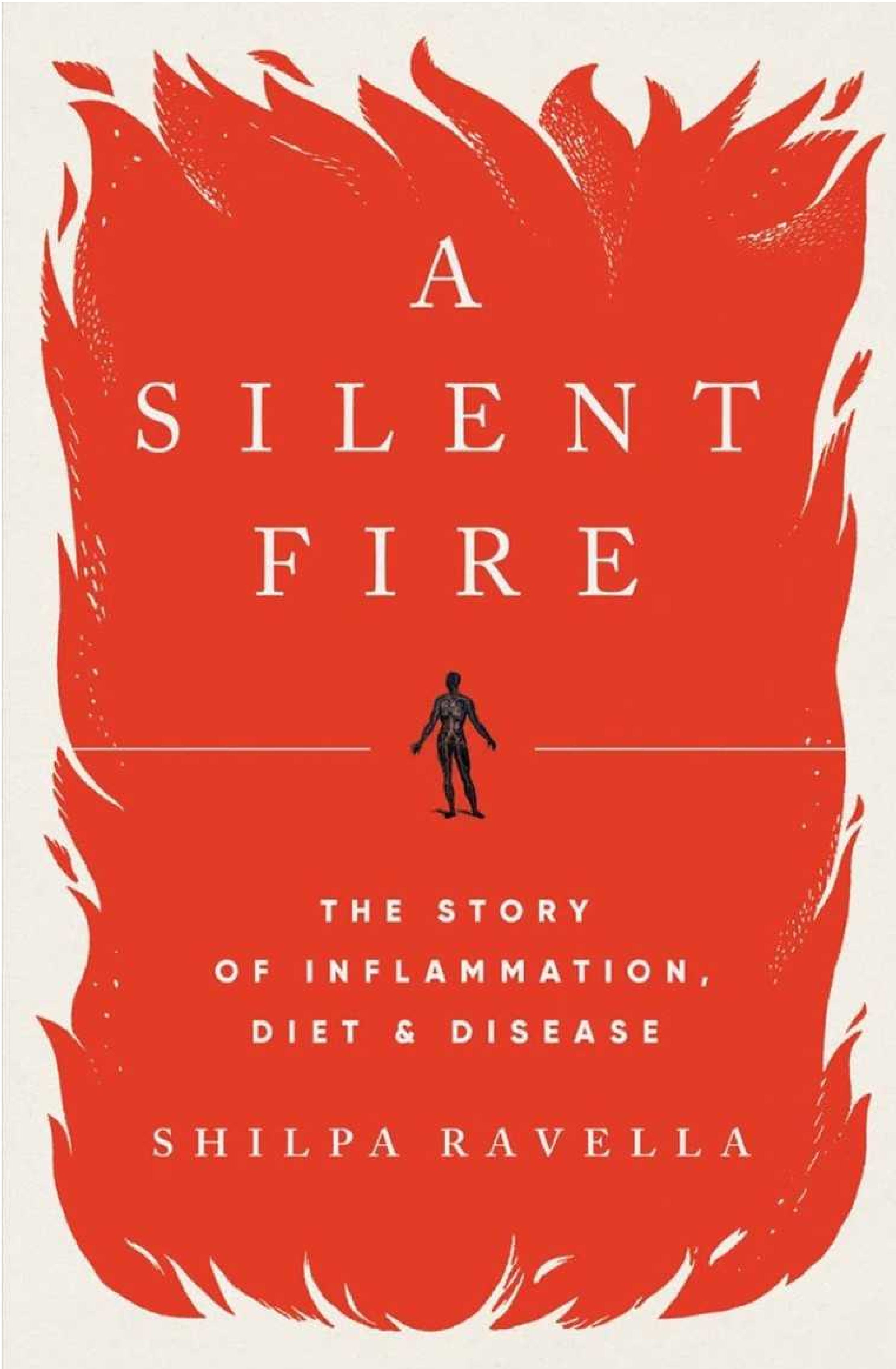
This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00887-4

- BOOK REVIEW
- 03 February 2023

When the immune system bites back, and biodiversity v. capitalism: Books in brief

Andrew Robinson reviews five of the best science picks.

- Andrew Robinson


A
SILENT
FIRE
THE STORY
OF INFLAMMATION,
DIET & DISEASE
SHILPA RAVELLA

A Silent Fire

Shilpa Ravella *W. W. Norton* (2022)

Inflammation of the body sounds threatening, but is a protective ancestral response to threats such as microorganisms, chemicals and traumas. It has “served us well through most of human history”, writes gastroenterologist Shilpa Ravella in her fascinating, case-oriented, story of inflammation, diet and disease. This is less true of ‘hidden inflammation’: a “silent fire” invisible to the person experiencing it and to tools for diagnosing inflammatory diseases, such as arthritis. Ravella focuses on controlling hidden inflammation through nutrition.

"You Are Not Expected to Understand This"

How 26 Lines of Code Changed the World

Edited by Torie Bosch

With illustrations by Kelly Chudler

“You Are Not Expected to Understand This”

Ed. Torie Bosch *Princeton Univ. Press* (2022)

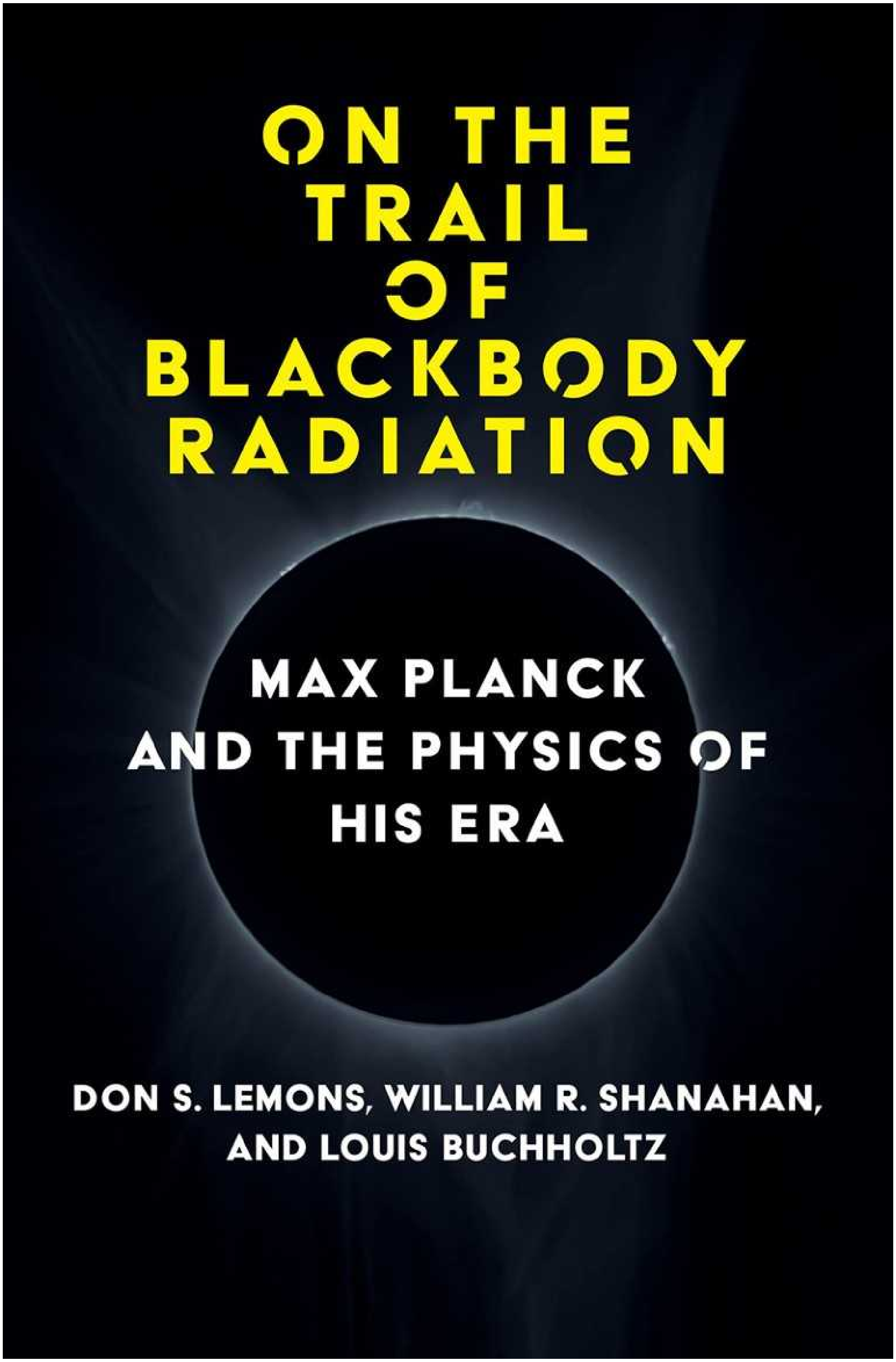
The title of this intriguingly human collection of articles about computer coding quotes a line from the Unix source code, written by a programmer in 1975. The code contained a frustrating bug that could not be fixed, so the software had to be entirely rewritten. Coding has always consisted of both wins and blunders, as discussed by the book’s 29 contributors, including programmers, technologists, historians, journalists and academics, covering subjects ranging from space exploration to human biases around race.


EVERYDAY LIFE
IN THE ICE AGE
Elle Clifford
Paul Bahn

Everyday Life in the Ice Age

Elle Clifford & Paul Bahn *Archaeopress* (2022)

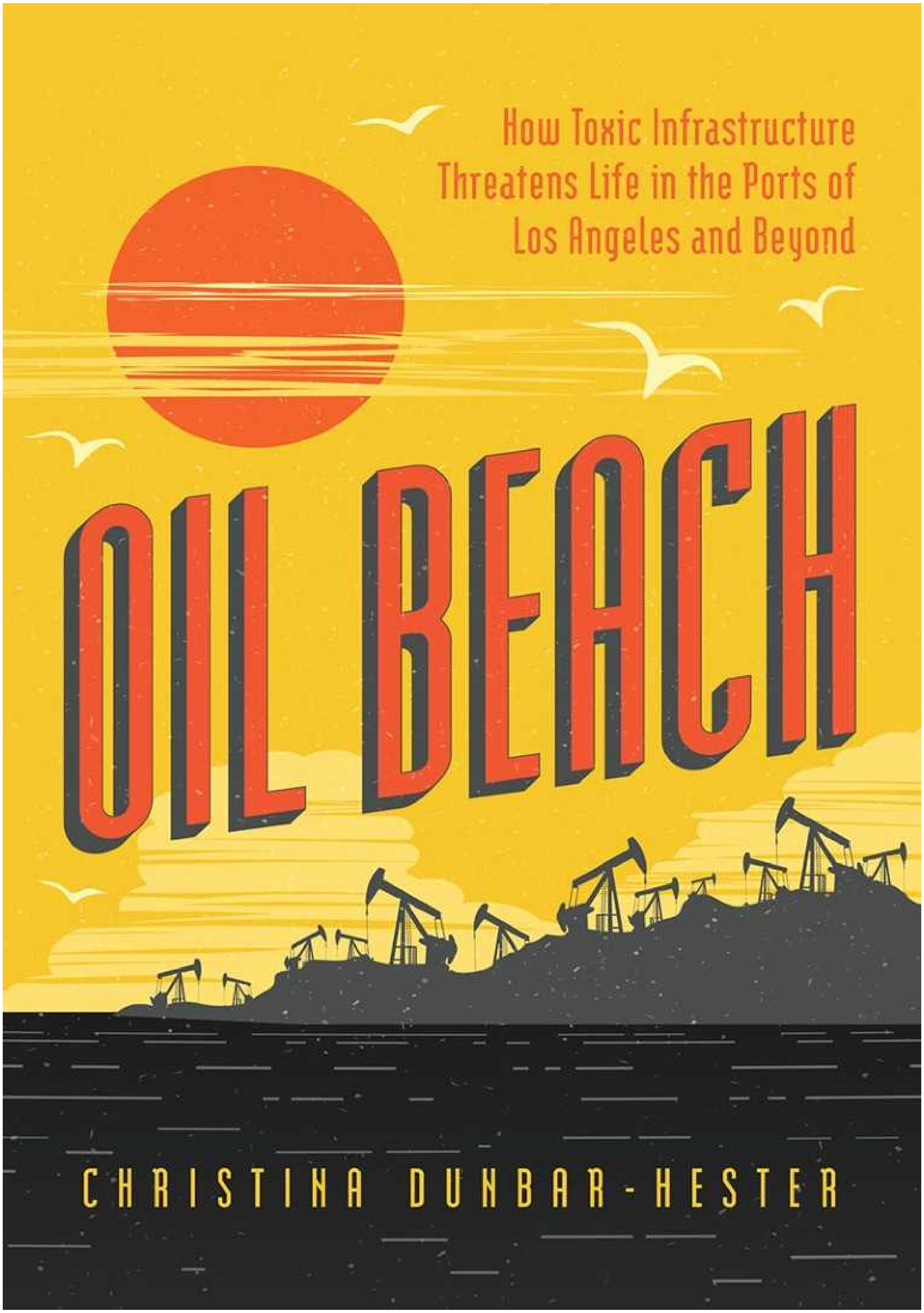
Most studies of the ice age 40,000–12,000 years ago focus, unexcitingly, on cataloguing tools. In this witty, richly illustrated look at European ice-age life — the first of its kind in English — social psychologist Elle Clifford and archaeologist Paul Bahn try to decipher tool use. Inevitably, swathes are speculative. Fewer than 200 human skeletons and fragments from the time survive; to supplement archaeological evidence, the authors use data from hunter-gatherers and knowledge of basic human behaviour, animals and plants.


ON THE TRAIL OF BLACKBODY RADIATION
MAX PLANCK AND THE PHYSICS OF HIS ERA
DON S. LEMONS, WILLIAM R. SHANAHAN, AND LOUIS BUCHHOLTZ

On the Trail of Blackbody Radiation

Don S. Lemons *et al*. *MIT Press* (2022)

Black-body radiation — emitted and absorbed by non-reflective bodies in thermal equilibrium — was named by Gustav Kirchhoff in 1862. But he and others were perplexed by calculations suggesting it should be infinite at high frequencies. This unavoidably mathematical history by three physicists follows the trail from Kirchhoff to Max Planck — who in 1900 explained that the radiation could change its energy only in minimal increments proportional to the wave's frequency — and Albert Einstein's quantum theory of radiation in 1917.


How Toxic Infrastructure
Threatens Life in the Ports of
Los Angeles and Beyond
OIL BEACH
CHRISTINA DUNBAR-HESTER

Oil Beach

Christina Dunbar-Hester *Univ. Chicago Press* (2023)

California has long been a byword for the clash between the environment and capitalism. Science and technology studies scholar Christina Dunbar-Hester investigates Long Beach near Los Angeles: a hub for petroleum shipping and refining, and a world-leading container-shipping port, but also a wetland area full of wildlife. Focusing on cetaceans, bananas, sea birds and otters, her informed interdisciplinary study considers how to avoid toxic tension between biodiversity and capital accumulation.

Nature **615**, 788 (2023)

doi: https://doi.org/10.1038/d41586-023-00333-5

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00333-5

Opinion

- **The Cape Town Statement on fairness, equity and diversity in research** [24 Mar 2023]
  • The benefits of scientific collaboration are too often skewed towards wealthier countries. Bioethicists and others present guidance on how stakeholders such as researchers can change this.
- **Why we need a new economics of water as a common good** [22 Mar 2023]
  • Anthropogenic pressures and climate change are altering water flows worldwide. Better understanding, new economic thinking and an international governance framework are needed to stave off catastrophe.

- COMMENT
- 24 March 2023

The Cape Town Statement on fairness, equity and diversity in research

The benefits of scientific collaboration are too often skewed towards wealthier countries. Bioethicists and others present guidance on how stakeholders such as researchers can change this.

- Lyn Horn ,
- Sandra Alba,
- Gowri Gopalakrishna,
- Sabine Kleinert,
- Francis Kombe,
- James V. Lavery &
- …
- Retha G. Visagie

Show authors



Libraries in many institutions in low- and middle-income countries cannot finance a broad range of journal subscriptions. Credit: Mamunur Rashid/NurPhoto/Getty

As awareness has grown about fraud and misconduct in science, the World Conferences on Research Integrity have become a leading forum for the discussion and study of ways to promote responsible behaviour in research. Since the first meeting in 2007, which was held in Lisbon, the events have helped to establish an academic field focused on research integrity.

Meetings have typically concentrated on issues such as research misconduct, responsible behaviour around data collection, analysis, authorship and publication, and the importance of reproducibility. But last May, attendees to the 7th World Conference on Research Integrity, held in Cape Town, South Africa, took a significant step. They added the myriad ways in which research programmes and practices disadvantage those living in low- and middle-income countries (LMICs) to the suite of issues that threaten the integrity of science.

Three of the six previous world conferences have led to the publication of guidelines or principles. We are part of a working group (including bioethicists, researchers, institutional leaders and journal editors) that now presents the Cape Town Statement on Fostering Research Integrity Through the Promotion of Fairness, Equity, and Diversity. (See Supplementary information for a list of working-group members.)

This statement entails 20 recommendations, drawn from discussions involving around 300 people from an estimated 50 countries, including 16 African nations and 5 South American ones. The discussions were held over 18 months — before, during and after the Cape Town conference — the theme for which was ‘fostering research integrity in an unequal world’.

The Cape Town Statement is essentially a call to action that we hope will help to turn the global conversation on inequity and unfairness in research into changes in practice by all stakeholders. Here, we lay out the motivation for the statement, and its broad goals.

Layers of injustice

The reasons for the statement are clear. Much too often, researchers and institutions from high-income countries reap greater benefits from global collaborations than do LMIC collaborators — whether in relation to numbers of papers published, authorship, career progression, setting priorities for research or the ownership of samples and data[1].



Nature addresses helicopter research and ethics dumping

As an indication of this, a look at the authorship of papers about COVID-19 from the 10 top medical and global-health journals (according to impact ratings), containing content related to Africa or any African country, and published during the first 9 months of 2020, reveals that 66% of the authors were not from Africa[2]. One in five articles had no author from Africa at all. What's more, of those papers with African authorship, 59% of first authors and 81% of last authors were not from Africa, and only 14% of papers had both an African first and last author.

Often, what happens is that after securing a grant for a project, a research team from a high-income country looks for local researchers in the low- or middle-income country of interest to collaborate with. Local researchers might be offered some grant money and co-authorship on a paper (usually with their name appearing in the middle of the list). Invariably, the lead research team conducts the analyses, with the local researchers only reviewing manuscripts, often to ensure that they are culturally and politically acceptable[3].

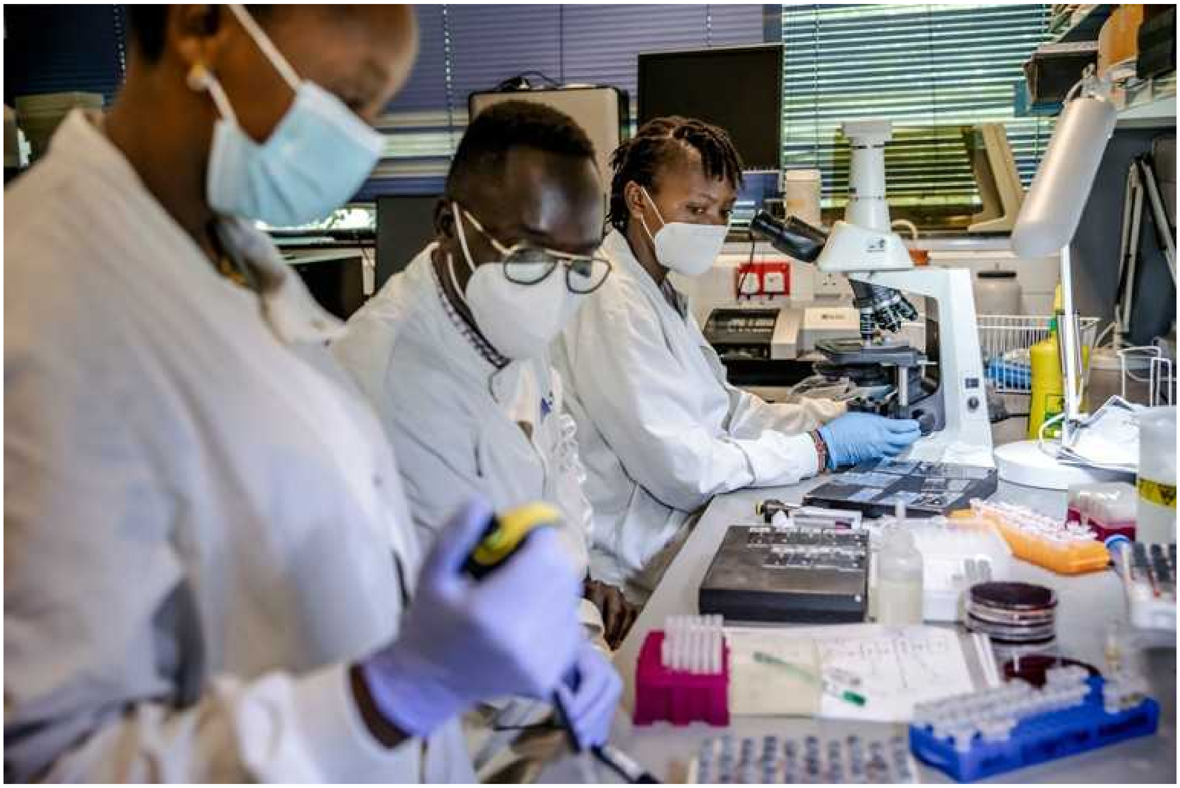
Even the push towards openness and transparency in science publishing — which many have argued is a way to foster greater integrity in research — has created more barriers for investigators in low-resource environments.

Sharing data, for example, requires having enough institutional infrastructure and resources to first curate, manage, store and (in the case of data relating to people) encrypt the data — and to deal with requests to access them. Also, the pressure placed on researchers of LMICs by high-income-country funders to share their data as quickly as possible frequently relegates them to the role of data collectors for better-resourced teams. With enough time, all sorts of locally relevant questions that were not part of the original project could be investigated by local researchers. But, well-resourced investigators in high-income countries — who were not part of the original project — are often better placed to conduct secondary analyses.

Unforeseen difficulties are arising around publishing, too. Currently, the costs to publish an article in gold open-access journals (which typically range from US$500–$3,000) are prohibitive for most researchers and institutions in LMICs. The University of Cape Town, for example, which produces around 3,300 articles each year, has an annual budget of $180,000 for article-processing costs. This covers only about 120 articles per year.

Because of this, researchers in these countries frequently publish their papers in subscription-based journals. But scientists working in similar contexts can't access such journals because the libraries in their institutions are unable to finance subscriptions to a wide range of journals. All this makes it even harder for researchers to build on locally relevant science.

Such imbalances in global research collaborations — which stem from a complex mix of racial discrimination, systemic bias and major disparities in funding and resources — impact the integrity of research in numerous ways.



Scientists in Kenya work on a malaria vaccine at the KEMRI-Wellcome trust laboratories in Kilifi.Credit: Luis Tato/Guardian/eyevine

As recently as 2019, a group of students and their professor at a university in South Africa drew racist conclusions from their study of the cognitive abilities of Coloured South African women (Coloured is a recognized racial classification in South Africa). Their findings have since been debunked, and an investigation concluded that there was no deliberate intention to cause harm. A lack of diversity among the researchers, and possibly among the reviewers and editors, might have contributed to the publication of this research, which has now been retracted by the journal involved.

Power imbalances also skew research priorities, with investigators pursuing goals that frequently overlook the needs of local people.

Improve fairness and equity

The Cape Town Statement is not the first guideline on research fairness and equity, particularly in collaborations. Indeed, various documents informed

our discussions, including the guiding principles of the Commission for Research Partnerships with Developing Countries (KFPE), which focuses on collaborations involving Swiss institutions; the Global Code of Conduct for Research in Resource-Poor Settings, a resource for those striving to ensure that science is carried out ethically in lower-income settings; and the BRIDGE guidelines, which aim to foster fairness and integrity in global-health epidemiology.

Furthermore, a tool for evaluating practices, called the Research Fairness Initiative, has already been developed by the Council on Health Research for Development, an international non-governmental organization that aims to support health research, particularly in LMICs. By providing questionnaires and guidance, the Research Fairness Initiative enables institutions, individual researchers and funders to evaluate their current practices, and if necessary, improve them.

The Cape Town Statement differs from these other guidelines and tools, however, in that it recognizes that unfair practices can harm the integrity of all research, no matter the discipline or context. Specifically, it focuses on the following four broad actions.

Increase diversity and inclusivity

More funders from high-income countries must include diversity stipulations in their calls for grant applicants. In 2020, for example, the second European and Developing Countries Clinical Trials Partnership programme (EDCTP2) in conjunction with the UK Department of Health and Social Care, asked applicants to apply for funding for projects specifically aimed at “addressing gender and diversity gaps in clinical research capacity” in sub-Saharan Africa. (The EDCTP is a partnership between countries in Europe and sub-Saharan Africa, supported by the European Union.)

Researchers in high-income countries must also work harder to collaborate in more meaningful ways with people from different disciplinary, geographical and cultural backgrounds. One way to do this is to establish long-term relationships with people that extend beyond the life of a single project.



South Africa’s San people issue ethics code to scientists

Research institutions — be they universities, non-governmental organizations, or national or transnational science councils — should develop and implement policies, structures and processes that support and promote diversity and inclusivity in research.

The London School of Hygiene and Tropical Medicine has been a major player in global-health research for many decades. Since 2019, a volunteer group of staff and students at the university has been trying to address the fact that the organization’s members have often held unacknowledged positions of advantage through its Decolonising Global Health initiative. So far, this has involved various undertakings, including a series of educational lectures.

Encourage fair practice in research

All stakeholders, from researchers, institutions and funders, to journal editors and publishers, must take steps to ensure that they are not exacerbating power imbalances in research collaborations, but instead helping to remove them.

Funders from high-income countries should discourage parachute or helicopter research — in which well-resourced researchers conduct studies in lower-income settings or with groups who have been marginalized historically, but fail to involve local researchers or communities in all stages of the research. Funders can do this by including diversity stipulations in their calls for grant applicants, but also by funding local researchers directly.

Although many need to follow suit, some funders are making progress on this front. Currently, about 60% of the grants awarded by the EDCTP go directly to institutions in sub-Saharan Africa. Similarly, in 2021 the US National Institutes of Health (NIH), launched a $74.5 million five-year project called Data Science for Health Discovery and Innovation in Africa (DS-I Africa). This is being led by African scientists, and is creating a pan-African data network designed to address African research priorities[4].

As well as requiring that researchers from LMICs lead collaborations, funders (and research institutions) should insist that projects involving multiple countries begin with a period of discussion involving all potential stakeholders. Before any research is conducted, the roles of each team member, and how they will receive recognition, should be defined and agreed on. Moreover, requiring that projects be fully and transparently budgeted, with the costs of maintaining already existing infrastructure included, would help to ensure that better resourced institutions shoulder a fairer share of project costs.

Publishers and journal editors must question submissions from authors if data have been collected in a low- or middle-income country, but the lead and collaborating authors are from high-income countries.

Some are already taking steps in this direction. *The Lancet* has started rejecting papers that are submitted by researchers from outside Africa, with data collected from Africa, but with no mention or acknowledgement of a single African collaborator. Similarly, Nature journals now encourage authors to make various disclosures on inclusion and ethics when submitting manuscripts.

The obstacles that make it harder for researchers working in low-resource settings to participate in ‘open science’ need to be identified and addressed

by publishers, and other national and global stakeholders, such as science councils and funders.

Wherever possible, funders should allow data collected by researchers in LMICs to be embargoed for two years, for example, to give investigators time to conduct secondary analyses and to share their data with collaborators at their discretion. Meanwhile, journals and publishers should adjust article processing costs for authors in low-resourced regions.

Over the past 20 years, more than 200 publishers have partnered with Research4Life, a platform established in 2002 dedicated to making some peer-reviewed content available to students and researchers in LMICs. But many of the countries that are major contributors to research, such as South Africa, Brazil, Argentina and India, do not meet the criteria for accessing knowledge through this platform, or for fee waivers or reductions in cost for open-access publishing.

Provide infrastructure

Researchers in LMICs are often disadvantaged because their institutions have underdeveloped research management and financial systems. With scant or no assistance from lawyers, administrative assistants, financial managers and project-management staff, they struggle to meet the 'due diligence' requirements of some funders in high-income countries. (Frequently this involves answering hundreds of questions in multi-page documents about institutional and research governance processes and policies.)

Both funders and collaborating institutions must take steps to enable the development of research-support systems in LMICs. This could mean paying for computing infrastructure, mentorship programmes, open-access publishing, or the training and salaries of project and financial managers, for example.



A volunteer has blood withdrawn for testing as part of an HIV vaccine trial in Bangkok.Credit: Paula Bronstein/Getty

Several controversies in research ethics, which occurred as a result of HIV research projects undergoing ethical scrutiny during the height of the HIV epidemic in the 1980s and early 1990s, prompted the Fogarty International Center at the NIH to launch the International Bioethics Education and Career Development Award in 1999. The aim was to ensure that there were strong research ethics committees in LMIC institutions, with adequately trained members, to review studies and meet international ethics and regulatory standards required by US funders. Thanks to this and subsequent efforts, hundreds of people in Africa and Asia have been trained in research ethics and continue to serve on review boards.

Many more initiatives like this — that invest in infrastructure and training over the long term — are needed.

The governments of LMICs also need to recognize the value of funding research, both to address locally relevant priorities and to reduce their nations' reliance on funders from high-income countries. Matched funding

schemes could help, whereby governments commit to give institutions the same amount of funds as those obtained from other sources for nationally identified high-priority research. Launched in 2015, the Science Granting Councils Initiative aims to strengthen the management of research grant funding in 17 countries throughout Africa. Achieving its goal of bringing more support for and control of scientific research into the continent will require governments of African countries to prioritize research funding.

Recognize Indigenous knowledge

During the 2000s, researchers from the United Kingdom and other high-income countries obtained blood samples from people of the San community in Namibia for genetic research without always adequately explaining what those samples would be used for, or reaching any benefit-sharing agreements with the community.

The San people have since developed their own code of ethics — a value-based set of principles that researchers must adhere to before trying to obtain samples or information from them (see go.nature.com/3yz7ash). In principle, this code could be used by researchers working with other Indigenous communities, if codes specific to a particular community don't yet exist.

Ensuring that community members or knowledge-holders, who might not have formal qualifications, are included in research teams — with their contributions being adequately valued — is another way in which local knowledge can be incorporated equitably[5].

Global equity for global problems

Many of the greatest challenges facing humanity — climate change among them — are disproportionately affecting people living in LMICs. And many of these challenges have arisen largely because of a long history of colonial exploitation and inequitable use of Earth's resources. Yet last year, reports of systemic bias and the contributions of researchers from LMICs being sidelined were made, even in the Intergovernmental Panel on Climate Change.

Unfairness, inequity and a lack of diversity must no longer prevent the global research enterprise from maximizing scientific integrity and from realizing the ultimate societal value and benefits of research.

Nature **615**, 790-793 (2023)

doi: https://doi.org/10.1038/d41586-023-00855-y

References

1. Parker, M. & Kingori, P. *PLoS ONE* **11**, e0163579 (2016).

2. Naidoo, A. V., Hodkinson, P., Lai King, L. & Wallis, L. A. *BMJ Glob. Health* **6**, e004612 (2021).

3. Odjidja, E. N. *Lancet* **9**, E1365 (2021).

4. Haberer, J. E. & Boum II, Y. *N. Engl. J. Med.* **388**, 387–390 (2023).

5. Schroeder, D., Chennells, R., Louw, C., Snyders, L. & Hodges, T. *Camb. Q. Healthc. Ethics* **29**, 285–301 (2020).

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00855-y

- COMMENT
- 22 March 2023

Why we need a new economics of water as a common good

Anthropogenic pressures and climate change are altering water flows worldwide. Better understanding, new economic thinking and an international governance framework are needed to stave off catastrophe.

- Johan Rockström ,
- Mariana Mazzucato,
- Lauren Seaby Andersen,
- Simon Felix Fahrländer &
- …
- Dieter Gerten

Show authors



Destruction of the Amazon rainforest adversely affects rainfall in Brazil's downwind neighbours. Credit: Maxime Aliaga/Nature Picture Library

Water is the lifeblood of our planet — essential for keeping humans and every plant and animal alive. It helps to circulate carbon and nutrients in the air and in soils, and regulates climate. For millennia, Earth's water cycle has provided reliable supplies and sustained conditions conducive to human development. Yet anthropogenic pressures are now pushing the cycle out of balance, threatening to undermine the reliability of rainfall itself.

The impacts are already being felt across the world — in devastating floods, such as those in Pakistan last year that killed 1,500 people and affected two-thirds of the country's districts, and in severe droughts such as the five failed rainy seasons in a row that have brought more than 20 million people to the point of starvation in the Horn of Africa. Meanwhile, more than 2 billion people still lack access to safe drinking water, one child dies every 17 seconds from waterborne diseases, and 3 billion people face food insecurity owing to water scarcity — numbers that could grow with the global population unless water provision improves (see go.nature.com/3jkgtry).

Water managers have always had to deal with natural variability, building larger reservoirs and tapping aquifers to fight scarcity, for example. But current challenges and trends in the rest of this century demand a completely different approach: a radical shake-up in how water is governed, managed and valued, from local to global scales, including a re-evaluation of human water needs (see Supplementary information, Box S1).

Today, the sector concentrates on flows of 'blue' fresh water — liquid that runs off the land and is stored in rivers, lakes, reservoirs and underground aquifers. Utilities capture and extract this water locally for drinking and sanitation, agricultural irrigation and industry. They assume it will be continually replenished, naturally, within historical ranges. In many places, that premise already no longer holds.



As the UN meets, make water central to climate action

Each 1 °C of global warming increases global mean precipitation by 1–3%, and it could rise by up to 12% by the end of the century compared with the period 1995–2014[1]. The impacts will be felt unevenly, with the frequency and severity of both floods and droughts rising. Deforestation, land degradation and infrastructure development are also altering precipitation patterns and affecting where water comes from and ends up[2]. Excessive extraction for irrigation and industry is aggravating water shortages in river

basins, from the Colorado in the United States and the Yangtze in China to the Murray–Darling in Australia.

To meet these growing challenges, water must be recast as a global common good. That means states establishing an obligation under international law to protect the global water cycle for all people and generations, and acknowledging that actions in one place have impacts in another — for instance, that deforestation in Brazil affects rainfall in Peru. It means assessing the role and economic value of not just blue fresh water, but also 'green' water that is held in the air, biomass and soils. And it means governments and the private sectors reformulating their roles and responsibilities, to develop objectives, policies and funds that can reshape markets and better manage global water supplies.

All these challenges must be discussed at the United Nations Water Conference in New York this week — the first such meeting in almost 50 years. Here, we highlight three areas in which research is badly needed to support discussions.

Understand all water flows within and between nations fully

Managing fresh water on a global scale means going beyond our current fixation on capturing blue water, which constitutes 35% of all fresh water on land, to also encompass green water, which makes up the remaining 65% (see Supplementary information, Fig. S1). Flows of moisture and vapour from land and vegetation are essential for regulating the water cycle and securing future rainfall, as well enabling carbon sequestration in soils and forests.

Globally, up to half of terrestrial precipitation originates from green water evaporated over land, with the rest from evaporation over the ocean[3]. Thus, landscape changes can alter water supplies in regions downwind, as well as changing local climates and streamflows. For example, deforestation in the Congo Basin lowers rainfall in neighbouring countries, and even across the Atlantic in the Amazon. Heavy irrigation of crops in India can boost the

streamflow of the Yangtze River in China, through moisture transported downwind[4].



Flash floods: why are more of them devastating the world's driest regions?

By analogy with watersheds on land, researchers refer to 'precipitationsheds' and 'evaporationsheds' in the atmosphere. Simply put, a precipitationshed is where rain comes from and an evaporationshed is where evaporation goes to. (Here, evaporation refers to total evaporation from the ocean and green water flows from land, including from soil and water bodies, as well as transpiration from vegetation.)

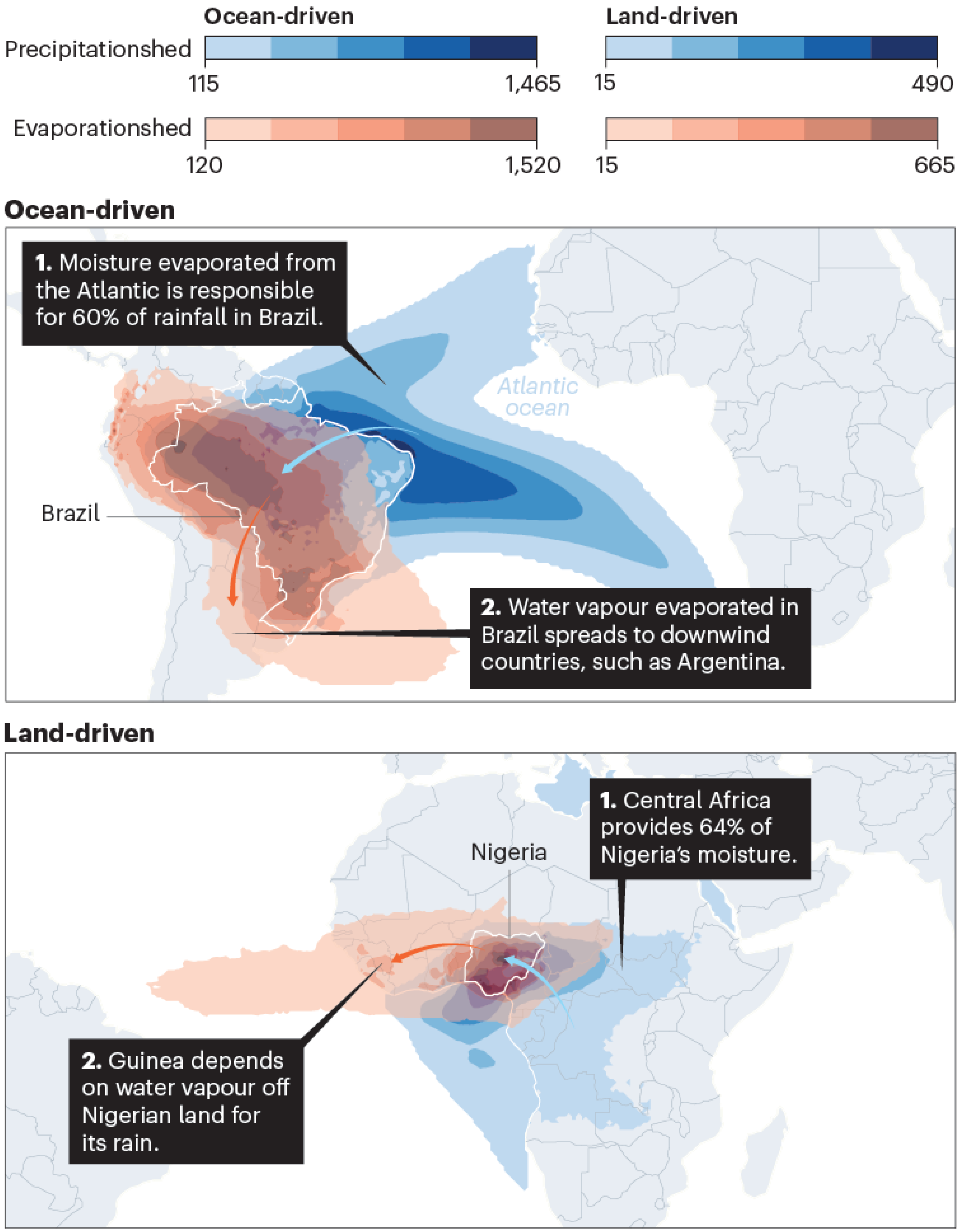
Researchers need to understand better how these processes interact and how atmospheric flows of water vapour connect different regions. A new view of interconnectivity is emerging, through combining meteorological databases (including on water vapour, humidity, wind speed and direction) and computer models that connect likely sources and sinks.

To illustrate, we used such data[3,5] to calculate volumes, ratios and flows of evaporation and precipitation in several regions (see 'Atmospheric watersheds', 'Terrestrial moisture flows across borders' and Supplementary information, Fig. S2). Generally, countries where prevailing winds blow from the ocean have a plentiful and consistent source of moisture and little dependence on other nations. Landlocked countries are more vulnerable to

natural variability and the practices of neighbours over which they have no control.

ATMOSPHERIC WATERSHEDS

The atmosphere harbours precipitationsheds (regions that act as sources of precipitation to a certain area) and evaporationsheds (regions that receive this area's evaporation, which falls as precipitation). In addition to evaporation from the ocean, the water cycle is driven by moisture from evapotranspiration from terrestrial land and vegetation, and land-cover changes in one country can affect another's rainfall.



Nature publications remain neutral with regard to contested jurisdictional claims in published maps.

©nature

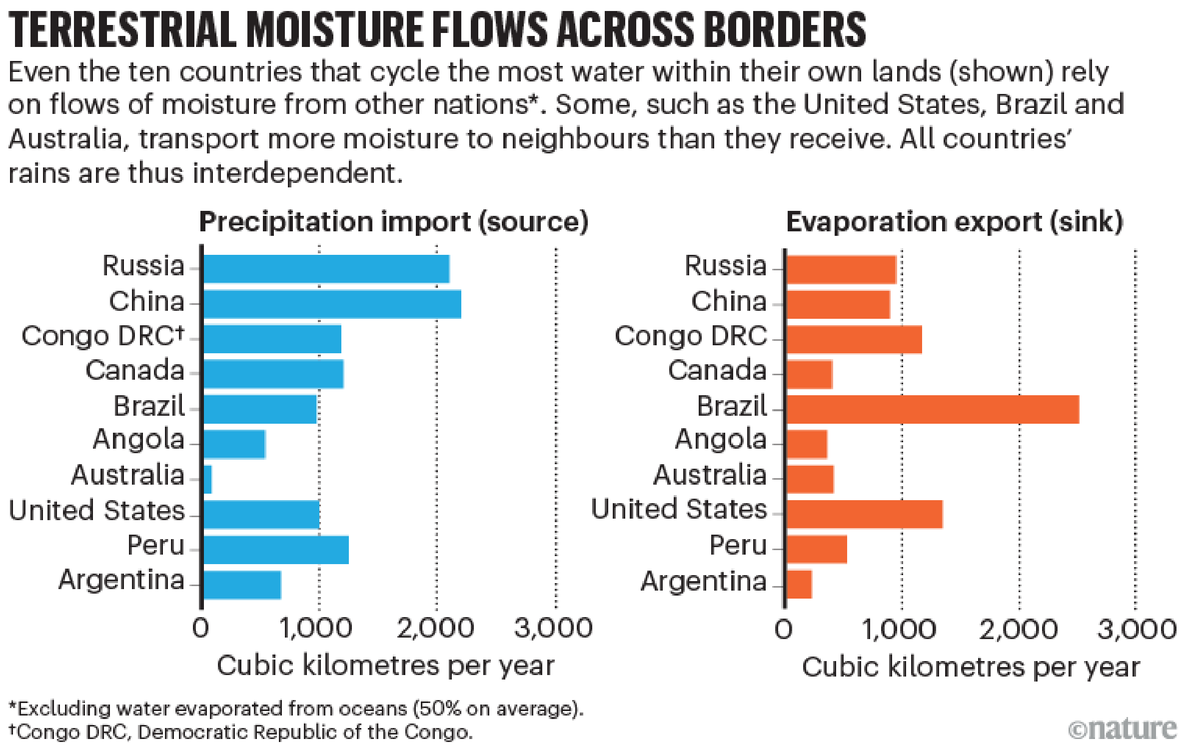
Source: Analysis by J. Rockström *et al.*

For example, Brazil is largely self-sufficient in green water and precipitation. We find that around 60% of its rainfall comes from moisture evaporated from the Atlantic, and 35% from moisture from Brazilian lands, including the Amazon rainforest (see Supplementary information, Fig. 2a). Much of this airborne moisture stays within the country, trapped by the High Andes. But Brazil also exports 25% of its green water to downwind countries, such as Argentina, Bolivia and Colombia. Rainfall in these nations will drop if deforestation in the Amazon continues[6], yet no political or institutional arrangements exist to address this dependency.

Rainfall patterns in sub-Saharan Africa, meanwhile, are tightly interwoven. Nigeria derives 64% of the moisture that precipitates its rainfall from within the continent; of this, 22% comes from within and 42% from outside its borders, predominantly from the Congo Basin. In turn, Nigerian land contributes 43% of the evaporated water driving rainfall in neighbouring countries such as Cameroon, Guinea and Ghana. All these countries' water supplies are thus at risk from deforestation in central Africa.

China, too, is heavily reliant (74%) on water evaporated from land for its precipitation. Of that, 44% comes from internally recycled moisture, and the rest from upwind neighbours, including India, Kazakhstan and Russia. Moisture from Chinese land also has a large role in rainfall across Central Asia and the Tibetan Plateau.

Moreover, no country acquires over half of its moisture from within its own boundaries, implying that even the largest countries rely on evaporation from other areas to sustain their precipitation. Even Russia, the most self-reliant in rainfall and with 45% of its moisture recycled internally (see Supplementary information, Fig. S2a), is still heavily dependent on neighbouring countries (20%) and the ocean (35%).



Source: Analysis by J. Rockström *et al*. See also Fig. S2b in the supplementary information.

This striking view of interdependence surpasses existing transboundary issues around rivers, lakes and groundwater, which are the usual focus of water governance and disputes. For example, the Grand Ethiopian Renaissance dam on the Blue Nile has an impact on supplies to Sudan and Egypt downriver. Researchers need to study how rifts between countries might grow once inter-reliance is better understood.

To inform policies, scientists need to assess water stocks and flows of green and blue water, locally and globally, using satellites, big data and Earth-system models. Researchers need to know where and through which processes global change is shifting freshwater cycles and supply. The impacts and costs of extreme events, such as parching of soils and extremes of river flow, need to be studied in the context of precipitationsheds and evaporationsheds.

Hydrologists, economists and political scientists will need to set budgets for green and blue water across scales, while keeping the sources and patterns of fresh water within ranges typical of the past 12,000 years during which

human civilizations evolved (the Holocene epoch). However, recent analyses suggest that features such as soil moisture are already deviating from historical ranges in some places, being either wetter or drier[7].

Rethink how water is valued and who 'owns' it

Treating water as a collective resource requires rethinking its economics. Currently, blue water is managed and regulated largely as a public good for drinking and sanitation. Yet public ownership undervalues water, in that one person's access does not limit another's, even though water is a finite resource. This promotes excessive, unsustainable and inequitable use. And it discourages private investment. In 2015, private-sector investment in water globally accounted for less than 5% of the total funds allocated to telecommunications, energy, transport and other basic services[8].



Degrowth can work — here's how science can help

By contrast, green water is given no economic value, despite the fact that it drives economic development, stabilizes climate change and secures precipitation. It can be public, private or a common good, depending on where it is.

To manage both blue and green water as a global common good, governments need to reshape water markets — not simply fix them when they fail. Governments must monitor soil moisture and vapour flows, and set policies that value these flows as natural capital. Water governance and management need to span all scales, connecting local watersheds, river basins, precipitation- and evaporationsheds, and eventually the globe.

To bring in businesses and investments, economists need to value water as an asset that generates functions and services for human well-being. This could follow, for example, the framework established in the Dasgupta Review on the economics of biodiversity, published by the UK government in 2021 (see go.nature.com/2om5hho), which sets value on natural capital and manages natural assets within a sustainability framework. Researchers must evaluate the amount of green water needed to sustain biodiversity and carbon sinks in ecosystems. And they must assess the 'social cost of water' (akin to the 'social cost of carbon'), which considers the costs to society of loss and damage caused by water extremes and not meeting the basic provision of water for human needs.



Access to clean drinking water remains a global problem, one that could worsen as the world's population continues to rise.Credit: Ashraf

Shazly/AFP/Getty

Whenever private companies benefit from public subsidies, guarantees, loans, bailouts and procurements, governments could attach conditionalities to contracts to maximize public benefits. For instance, the 1996 amendments to the Safe Drinking Water Act in the United States promoted equitable access to water by creating the Drinking Water State Revolving Fund to subsidize companies that provide water for disadvantaged communities. Similarly, the 2022 US CHIPS and Science Act contractually obliges funding recipients to maximize efficiency with regard to water, waste and electricity.

New forms of public–private arrangements, including permits, property rights and procurements, should be developed to counteract the rent-seeking and value-extractive behaviour that has plagued some national water sectors. In England, for example, since the privatization of the water industry in 1989, £72 billion (US$88 billion) has been paid out to shareholders as dividends, while outdated infrastructure has left the water system riddled with leaks and sewage discharges.

Some preliminary work to reshape the economics of water has begun: two of us (M.M. and J.R.) are leaders on the independent Global Commission on the Economics of Water (watercommission.org), which was launched in May 2022 at the World Economic Forum in Davos, Switzerland. The group is assessing impacts on the global hydrological cycle from climate and environmental change, as well as country interdependencies and the international cooperation needed to treat water as a global common good. A call to action (see go.nature.com/3zxnw54) and a first review report (see go.nature.com/3twxsok) were released the week before the UN 2023 Water Conference.

Start locally and build globally

Effective management of water as a global common good starts locally. National governments, cities and regions need to define goal-driven 'missions' that add up globally. For example, nations might pledge to ensure that the supply of green and blue water in the hydrological cycle inside their

borders remains within a manageable range, as defined by safe planetary limits or boundaries[9]. Targets and strategies must be designed to initiate coordination, finance and innovations[10].

For example, the European Union's Water Framework Directive has, since 2000, required the EU member states to develop river-basin management plans jointly with the public. Obligations are reviewed every six years, and non-compliance brings legal sanctions. Although progress has been made, more-coordinated efforts and monitoring would help to realize the directive's full potential[11].

All sectors must be involved. Food production, for example, accounts for around 75% of freshwater consumption globally, with India the largest consumer. India might, for instance, focus on ensuring continuity of food supply without imposing pressures on national use of green and blue water. Production and consumption processes should be redesigned to minimize water waste and maximize water sharing.

In Australia, the national science agency, CSIRO, is working to reduce the economic impacts of the country's droughts by 30% this decade, by making climate data accessible to farmers to enable them to make informed water-use decisions. Other countries, such as Kenya, are exploring 'green water credits' that reward upstream water management beneficial to downstream areas[12].

Cooperation and exchange of knowledge will be crucial to join up local and global strategies. As with greenhouse-gas accounting and the Sustainable Development Goals, the UN and other bodies will need to develop mechanisms for overseeing the planet's water resources; discussions on how to do that must start this week in New York City. None of what we set out here will be easy. But the future of Earth's bloodstream is at stake.

Nature **615**, 794-797 (2023)

doi: https://doi.org/10.1038/d41586-023-00800-z

References

1. IPCC. *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Masson-Delmotte, V. *et al.*) (Cambridge Univ. Press, 2021).
2. Staal, A. *et al. Environ. Res. Lett.* **15**, 044024 (2020).
3. Tuinenburg, O. A., Theeuwen, J. J. E. & Staal, A. *Earth Syst. Sci. Data* **12**, 3177–3188 (2020).
4. Wang-Erlandsson, L. *et al. Hydrol. Earth Syst. Sci.* **22**, 4311–4328 (2018).
5. Hersbach, H. *et al. Q. J. R. Meteorol. Soc.* **146**, 1999–2049 (2020).
6. Wunderling, N. *et al. Proc. Natl Acad. Sci. USA* **119**, e2120777119 (2022).
7. Wang-Erlandsson, L. *et al. Nature Rev. Earth Environ.* **3**, 380–392 (2022).
8. Kolker, J. E., Kingdom, B., Trémolet, S., Winpenny, J. & Cardone, R. *Financing Options for the 2030 Water Agenda: Water Global Practice Knowledge Brief* (World Bank, 2016).
9. Zipper, S. C. *et al. Earth's Future* **8**, e2019EF001377 (2020).
10. Mazzucato, M. *Mission Economy: A Moonshot Guide to Changing Capitalism* (Allen Lane, 2021).
11. Voulvoulis, N., Arpon, K. D. & Giakoumis, T. *Sci. Total Environ.* **575**, 358–366 (2017).
12. Dent, D. in *Soil as World Heritage* (ed. Dent, D.) 459–470 (Springer Netherlands, 2014).

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00800-z

| Section menu | Main menu |

Work

- **Pandemic productivity loss: how scientific institutions should support academic mothers** [28 Mar 2023]
  • Three years on, scientist mums implore universities, funding agencies and publishers to heed calls to account for COVID-19 disruptions.
- **Guardian of Ecuador's diverse — and vanishing — frog species** [27 Mar 2023]
  • Conservation biologist Andrea Terán-Valdez aims to protect endangered frogs in Ecuador by breeding populations, cataloguing new species and fighting industrial development.

- CAREER FEATURE
- 28 March 2023
- Correction 30 March 2023

Pandemic productivity loss: how scientific institutions should support academic mothers

Three years on, scientist mums implore universities, funding agencies and publishers to heed calls to account for COVID-19 disruptions.

- Nikki Forrester

Find a new job



Kimberly Griffin says that the COVID-19 pandemic continues to disrupt researchers' careers.Credit: Amy Laakso/CYC Teacher

The COVID-19 pandemic has had a disproportionate impact on mothers in academic science, bringing long-standing issues to the fore. A study done during the very early stages of the pandemic found that scientists with at least one child aged five or younger were spending 17% less time on research than they had previously[1]. Also, during this time, scientist mothers were significantly less likely to submit manuscripts and to meet deadlines than other researchers[2], which is known to lead to fewer professional opportunities and to many leaving the workforce altogether. "Moms leave work because the system is not designed to support them," says Wally Fulweiler, an ecosystems ecologist at Boston University in Massachusetts.

As pandemic restrictions eased, most parents and caregivers no longer had to oversee remote schooling, but many were still unable to travel because of restrictions or concerns about exposing their families to COVID-19. For

instance, Japan continued to limit the number of travellers allowed into the country until October 2022, and Fiji didn't lift the last of its travel restrictions until February 2023. Along with travel problems, many scientist mums still struggle to find the support needed to progress their careers. Universities throughout the world failed to acknowledge "that it is a 24-hour job to take care of kids at home", says Fernanda Staniscuaski, a biologist at the Federal University of Rio Grande do Sul in Porto Alegre, Brazil, and founder of Parent in Science, a campaign group started in 2016 to support researcher parents in Brazil. And such career breaks can lead to a vicious downward spiral of productivity: mothers publish less often, then earn less funding and in turn produce less, she says. "We need to break the cycle."



[The career cost of COVID-19 to female researchers, and how science should respond]

Ever since the pandemic took hold three years ago this month, scientist mothers have been asking academic institutions, publishers, scientific societies and funding agencies to address inequities and provide better support through policy changes and practical mechanisms. In 2021, Fulweiler and her colleagues set out specific recommendations for organizations around the world, such as providing supplemental funding for salary support, blending virtual and in-person conferences and re-examining tenure timelines[3]. But few of these recommendations have been implemented. And in January, Mothers in Science, an international non-

profit organization based in France, released an action plan for funding agencies that advocates for providing support for the impacts of COVID-19.

Nature talked to five mothers in academia about what institutions have done to make adjustments for pandemic disruptions and what other reinforcements parents, and especially mothers, need.

MEI LIN NEO: Improve family-friendly work policies

Founding member of Mothers in Science and a marine biologist and senior research fellow at the National University of Singapore.

At the start of the pandemic, my toddler was 15 months old. I struggled to do academic work and parenting simultaneously. I constantly felt as if I was drowning, with no means to surface for air. The pandemic also blurred the lines between spaces — our home, my workplace and our childcare centre converged into a single, tight space. One impact of this for me was a lack of energy, motivation and time to write research proposals, which continues to affect me now as I scramble to secure new funding.

My hope is that institutions will acknowledge that mothers are crucial in academia. For us to continue being effective in our workplaces, institutions must provide basic infrastructure to support us, such as lactation rooms. They should also implement family-friendly policies, such as allowing mothers to bring their kids to work when childcare is unavailable. The lack of these basic necessities reflects how little value institutions place on the contributions of parents.

When I first became a mother, my daughter's needs were my priority. I asked if she could be present in my meetings and tag along with me when I went to the lab. I was fortunate that most of my colleagues were supportive of my requests, which was surprising to me as I had never seen any other mothers do what I was doing in the workplace. So don't be afraid to ask, and to do what is best for you and your children.



Mei Lin Neo took her daughter to her laboratory when she could.Credit: Mei Lin Neo

WALLY FULWEILER: Provide subsidized support for childcare

Ecosystems ecologist at Boston University in Massachusetts.

The disproportionate amount of domestic care work that mums do also translates into the workplace: moms have to leave to pick up children from childcare, are not able to travel to conferences and have less time for their work. In science, thinking time is crucial, because you can’t write good papers, do data analysis or write proposals without it. During COVID-19 lockdowns, that time just vanished. But the loss of time and opportunities continued for those with young, unvaccinated kids even after lockdowns ended. For instance, childcare centres frequently closed because of COVID-

19 cases. Many scientists chose not to travel to conferences, give seminars or do fieldwork because of the risk of exposure for their laboratories, themselves and their families.

There was this big decline in the number of women, and moms in particular, submitting papers and proposals. It's not a mystery; a human can only do so many things.

When we were drafting a paper[3] outlining how institutions could support academic mothers, we had Zoom meetings during the day, with our kids crawling over us, or at night, when our kids were asleep. We brainstormed a range of ideas that we thought would be doable and helpful for mentors, university administrators, scientific societies, publishers and funding agencies. For example, they could provide relief from serving on committees, offer childcare and lactation support at meetings and invite mothers to write review articles and serve on editorial boards. We also recommended changes that didn't cost much to implement, such as asking supervisors to be flexible about timelines. But at the end of the day, if you really care about offering more support to scientist mothers, then you're going to have to put some money behind it.

Even little things are going to send a sign that everyone in leadership, from university administrators to funders to mentors, is thinking about this and wants moms to be here. This is a solvable problem; we just need to put the work in to solve it. Having universal, high-quality childcare is the number one change I'd like to see. Some options for universities include providing on-campus childcare, with a sliding scale for payment according to the parents' career stage, or funds to help parents pay for off-campus care. Childcare is essential infrastructure — parents can't work without it.

LOESKE KRUUK: Give caregivers unique publishing opportunities

Loeske Kruuk is a senior editor for Proceedings of the Royal Society B and an evolutionary ecologist at the University of Edinburgh, UK.

I was in Australia during pandemic lockdowns and there were a few months when I was home-schooling my children. I talked to other colleagues in the same situation and realized that we were all struggling with not having enough hours in the day.



Juggling scientific careers and family life

I found a paper[3] published by Fulweiler and her colleagues about supporting academic mothers during COVID-19 and beyond. There were a series of suggestions for what institutions could do, and one of the categories was for publishers. As an editor of *Proceedings of the Royal Society B*, I thought I could use my position at the journal to do something useful.

Last May, we published a special issue showcasing research in evolutionary biology from early-career academic mothers and care-givers. My two co-editors, Maurine Neiman and Sarah Brosnan, and I advertised on social media and sent messages asking people to send us abstracts and outline their experiences during the pandemic.

We received very positive responses both from our editor colleagues and from authors about the opportunity. The special issue was a collection of papers from early-career mothers who were doing top-quality science by any objective measure. The fact that they produced papers while dealing with the challenges of caring for their children during the pandemic makes their work all the more impressive.

One of the most valuable outcomes of the special issue was raising the profile of the challenges faced by academic mothers. I hope it was helpful for people to see that institutions are taking the problem seriously and want to do something about it. We need people to consciously include women in the publishing process as much as possible, by inviting them to submit papers, be reviewers and join editorial boards.



Efiba Vidda Senkyire Kwarteng promotes leadership for mums.Credit: Efiba Vidda Senkyire Kwarteng

EFIBA VIDDA SENKYIRE KWARTENG: Help mums secure leadership positions

Geomatics engineer at the Kwame Nkrumah University of Science and Technology in Kumasi, Ghana.

I have three kids, who are eight, four and one-and-a-half years old. Lockdown provided an opportunity for me to bond with them and get more involved in their education. We all shared a laptop. When it was time for them to join a Zoom class for school, I had to leave what I was doing to give them the laptop. I would often work at night to catch up on everything I missed during the day.

Once I could start travelling for work again, I did some training for a mining company to teach their workers how to use a geographic information system to collect and analyse data. I told them I had a baby, and they provided hotel rooms for me and my nanny during the first phase of the training. During the second phase, we camped at a mining site that didn’t have any microwaves or stoves in the rooms. I explained to them that I needed to be able to prepare food for my baby, and they quickly gave us access to a stove, fridge and everything else we needed. This company was doing its best, but other institutions can do more.

In universities, there should be mum-friendly facilities that give parents a place in which they can bond with their children and also work effectively. If I took my baby to my office, the moment she started crying it would disturb people in the other offices. The facilities could have a common play area and nearby offices designated for each mum. We do have a childcare centre for infants at my university. I took my second child there when she was about five months old because I couldn’t get any work done when I tried to take her to the office.



A call to create funding equity for researcher-mums

Along with good facilities, there should be more opportunities for mothers to move into leadership positions. We lead at home and we lead in the workplace, but many of us lose out on professional opportunities owing to the limits placed on our time. For instance, a more junior male colleague was given a leadership position that I was in line for because it required long hours, working at weekends and a heavier workload than I could take on. That's the situation we find ourselves in. If universities want to give these positions to mums, then they need to come up with new evaluation standards so that we don't have to neglect our families for our careers. It's not about lowering the standards so much as formulating them in a way that tells mothers that they are qualified and would fit those roles.

KIMBERLY GRIFFIN: Redefine productivity

Dean of the College of Education at the University of Maryland in College Park.

One of the most challenging aspects of academic work is that it's not a 40-hour-a-week job with a clear start and end time. There are no boundaries on how much you could work, or should work, to reach your goals. Parents and primary caregivers have real limitations on their time, but at the end of the

day, they’re being compared with people who don’t have those same limitations.

I navigated the pandemic with a child who’s now four years old. It was a time when we talked about people not being productive, and that seemed strange because it was my sense that people were working quite hard — they just weren’t working on things that academia rewards us for.



[The parenting penalties faced by scientist mothers]

In 2020, my colleague Leslie Gonzales, an associate professor of higher education at Michigan State University in East Lansing, and I published recommendations for how institutions could support faculty during and beyond COVID-19 (see [go.nature.com/3tv3jap]). Specifically, we suggested that committees and external reviewers apply different standards when evaluating productivity before and after lockdowns. For instance, expectations that faculty members who have yet to get tenure publish two papers per year should not apply for the years 2020 or 2021. They should also consider how the transition to online learning might have negatively affected student evaluations of their professors.

Productivity is often measured by the number of papers you published, grants you received and classes you taught. But those metrics miss all the emotional and invisible labour that faculty members engage in, such as

informal mentoring. Administrators and reviewers should invite faculty members to detail the emotional labour they provide and to document mentorship efforts, workshops and training they participated in.

As administrators, we need to think about how we can provide faculty members with the support they need to meet expectations. Are we providing access to caregiving? Are we providing funding to hire another graduate student? Are we providing more time by granting sabbaticals or releases from teaching courses or from significant service responsibilities?

At the University of Maryland, we’ve instituted a lot of new policies, particularly regarding tenure and promotion. We send external promotion reviewers a statement about the impacts of COVID-19, and candidates can also write a COVID-19 impact statement that is reviewed internally. The pandemic isn’t something that happened a couple of years ago for a few months — it is an ongoing disruption.

Nature **615**, 955-957 (2023)

doi: https://doi.org/10.1038/d41586-023-00888-3

These interviews have been edited for length and clarity.

Updates & Corrections

- **Correction 30 March 2023**: This Career feature erroneously described Mei Lin Neo as a co-founder of Mothers in Science. She is, in fact, a founding member.

References

1. Myers, K. R. *et al. Nature Hum. Behav.* **4**, 880–883 (2020).
2. Staniscuaski, F. *et al. Front. Psychol.* **12**, 663252 (2021).
3. Fulweiler, R. W. *et al. PLoS Biol.* **19**, e3001100 (2021).

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00888-3

- WHERE I WORK
- 27 March 2023

Guardian of Ecuador’s diverse — and vanishing — frog species

Conservation biologist Andrea Terán-Valdez aims to protect endangered frogs in Ecuador by breeding populations, cataloguing new species and fighting industrial development.

- Nikki Forrester



Andrea Terán-Valdez is a conservation biologist at the Jambatu Center for Amphibian Research and Conservation in San Rafael, Ecuador. Credit: Jaime Culebras

We are facing a global crisis. According to the International Union for Conservation of Nature, 41% of frog species globally and 57% of the frog species in Ecuador are endangered, owing primarily to chytrid fungus disease, habitat destruction and climate change. One of our emergency tools for protecting endangered frog species is to take frogs from the wild, breed them in the laboratory and release the offspring into their habitats.

My job involves working out how to do that. I also describe new frog species in an encyclopaedia of amphibians of Ecuador. So far, we have gathered information on 652 species. There are many more, and many are disappearing before we can describe them.

In 2021, one of my collaborators collected members of a new species of harlequin frog (*Atelopus* sp. nov.) from a recently logged forest. We're now trying to produce a new population. In this picture, I am measuring the eyes, fingers, legs and other features of this species. I record where and when the frogs are collected, and monitor how their habitat changes over time.

We have to know what species exist where, so we can protect them. In Junín, Ecuador, water quality will be affected if planned copper-mining operations are approved. My colleagues have found a critically endangered harlequin-frog species in the area, and a new frog species that lives in a waterfall. We fear that both species will go extinct if the mining goes ahead.

Along with community members and others, I presented a legal case to Ecuador's environment ministry to stop the mining, to protect these frog species. We won the case, but lost on appeal; now, I'm providing scientific information for a new filing.

Finding and describing new frog species is a powerful conservation tool because it compels governments to protect them. We don't want to lose any other species — we have already lost too many.

Nature **615**, 960 (2023)

doi: https://doi.org/10.1038/d41586-023-00889-2

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00889-2

Research

- **Fastest-ever calcium sensors broaden the potential of neuronal imaging** [15 Mar 2023]
  • Proteins have been developed that emit flashes of light in response to influxes of calcium ions into cells on millisecond timescales. Two sets of scientists discuss the legacy and future of these proteins.
- **Forming intracluster gas in a galaxy protocluster at a redshift of 2.16** [29 Mar 2023]
  • Analysis of observations from the Atacama Large Millimeter/submillimeter Array showed evidence of the thermal Sunyaev–Zeldovich effect in the direction of the Spiderweb protocluster at a redshift of 2.156.
- **Universal logic with encoded spin qubits in silicon** [06 Feb 2023]
  • In this alternative approach to quantum computation, the all-electrical operation of two qubits, each encoded in three physical solid-state spin qubits, realizes swap-based universal quantum logic in an extensible physical architecture.
- **Entwined African and Asian genetic roots of medieval peoples of the Swahili coast** [29 Mar 2023]
  • Analysis of ancient human DNA from the Swahili coast reveals that predominantly African female ancestors and Asian male ancestors formed families after around ad 1000 and lived in elite communities in coastal stone towns.
- **Diminishing benefits of urban living for children and adolescents' growth and development** [29 Mar 2023]
  • The advantage of living in cities compared with rural areas with respect to height and BMI in children and adolescents has generally become smaller globally from 1990 to 2020, except in sub-Saharan Africa.
- **Fast and sensitive GCaMP calcium indicators for imaging neural populations** [15 Mar 2023]
  • Using large-scale screening and structure-guided mutagenesis, fast and sensitive GCaMP sensors are developed and optimized with improved kinetics without compromising sensitivity or brightness.
- **Population dynamics of head-direction neurons during drift and reorientation** [22 Mar 2023]

• Mice maintain the memory of previous associations between head direction neurons and allocentric cues and this influences the internal head direction representation.

- **The menin inhibitor revumenib in KMT2A-rearranged or NPM1-mutant leukaemia** [15 Mar 2023]
  • Revumenib, a potent and selective oral inhibitor of the menin–KMT2A interaction, is associated with a low frequency of treatment-related adverse events and promising clinical activity in patients with relapsed or refractory acute leukaemia.
- **Whole-genome doubling drives oncogenic loss of chromatin segregation** [15 Mar 2023]
  • Whole-genome doubling induces the loss of segregation of chromatin compartments, and can lead to tumour-promoting epigenetic and transcriptional modifications.
- **Structural basis of mitochondrial membrane bending by the I–II–III2–IV2 supercomplex** [22 Mar 2023]
  • A supercomplex comprising all four respiratory chain components contributes to the induction of mitochondrial membrane curvature and tubulation of cristae.
- **Ultrafast structural changes direct the first molecular events of vision** [22 Mar 2023]
  • One picosecond after photoactivation, isomerized retinal pulls away from half of its numerous interactions with its binding pocket, and the excess of the photon energy is released through an anisotropic protein breathing motion in the direction of the extracellular space.

- NEWS & VIEWS FORUM
- 15 March 2023

Fastest-ever calcium sensors broaden the potential of neuronal imaging

Proteins have been developed that emit flashes of light in response to influxes of calcium ions into cells on millisecond timescales. Two sets of scientists discuss the legacy and future of these proteins.

- Michael B. Ryan ,
- Anne K. Churchland ,
- Yiyang Gong &
- …
- Casey Baker

Show authors

THE PAPER IN BRIEF

• Genetically encoded calcium indicators (GECIs) are designer proteins that react to intracellular calcium-ion (Ca^{2+})influxes — which occur during the activation of neurons — by emitting light.

• GECIs have enabled researchers to gain myriad insights into the workings of the brain, but until now the proteins have lacked the precision to record a fundamental unit of activity, known as the action potential, in single neurons.

• Writing in *Nature*, Zhang *et al.*[1] describe a GECI that marks a step towards solving this problem.

MICHAEL B. RYAN & ANNE K. CHURCHLAND: Lamps that illuminated the brain

Imagine rowing across still, dark waters on a summer night, when the water is suddenly transformed by the brilliant glow of countless jellyfish illuminating a hidden world of activity. In the 1960s, scientists from the Friday Harbor Laboratories in Washington collected nearly 10,000 specimens of *Aequorea* jellyfish by hand, with the goal of isolating the organic component responsible for this natural bioluminescence[2]. The scientists extracted and purified a calcium-sensitive protein that they named aequorin, together with an associated green fluorescent protein (GFP). Soon after, researchers loaded purified aequorin into living cells to track changes in Ca^{2+}, an ion that is central to neuronal activity[3]. This ability to measure neuronal activity by proxy — through flashes of light — revolutionized the field of neuroscience, and led to the identification of previously unknown dynamics in the brain.

By injecting aequorin into neurons, scientists discovered how changes in the levels of intracellular Ca^{2+}, resulting from neuronal activation, contribute to the process by which these cells release neurotransmitters to communicate with one another[4]. Several decades after aequorin was first isolated, another breakthrough came in the form of GECIs. Whereas aequorin and other calcium-sensitive dyes require manual loading into cells, GECIs are engineered to become part of the organism's DNA[5,6]. The most widely used GECIs, known as the GCaMP family[7], consist of a calcium-binding protein called calmodulin and a bound peptide (a short string of amino acids), fused to a modified form of GFP. When Ca^{2+} binds to calmodulin, this triggers a conformational change in the fusion protein, which causes GFP to emit light (Fig. 1a).

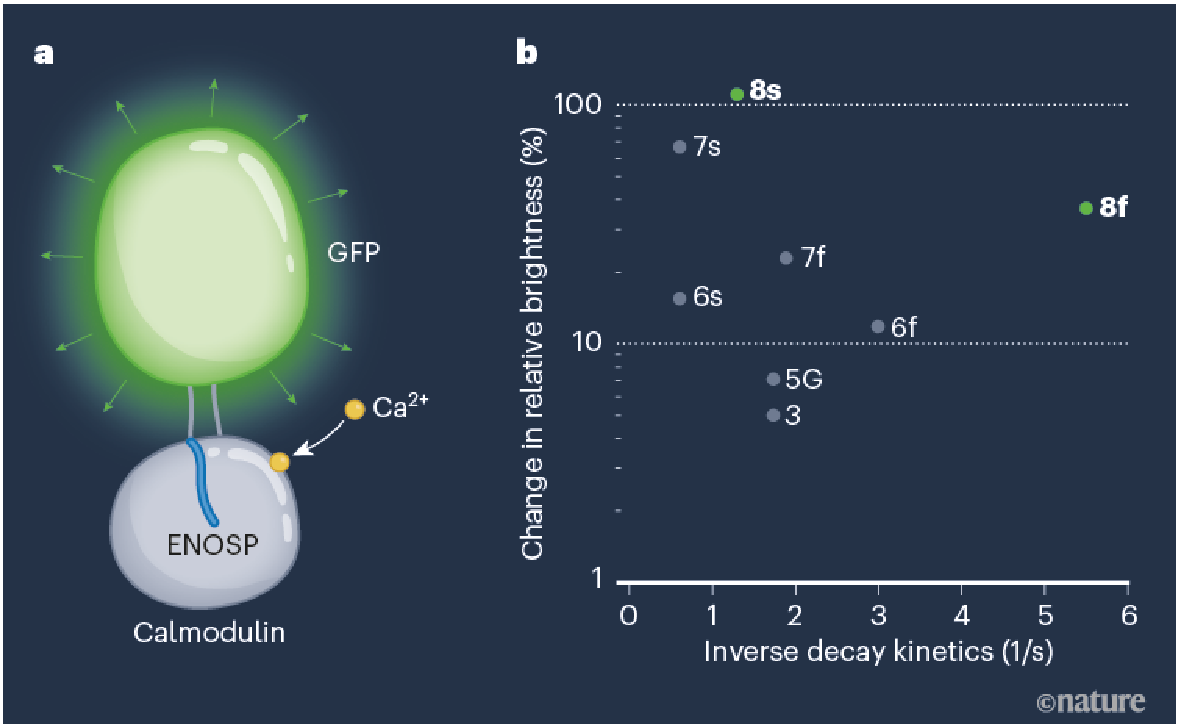
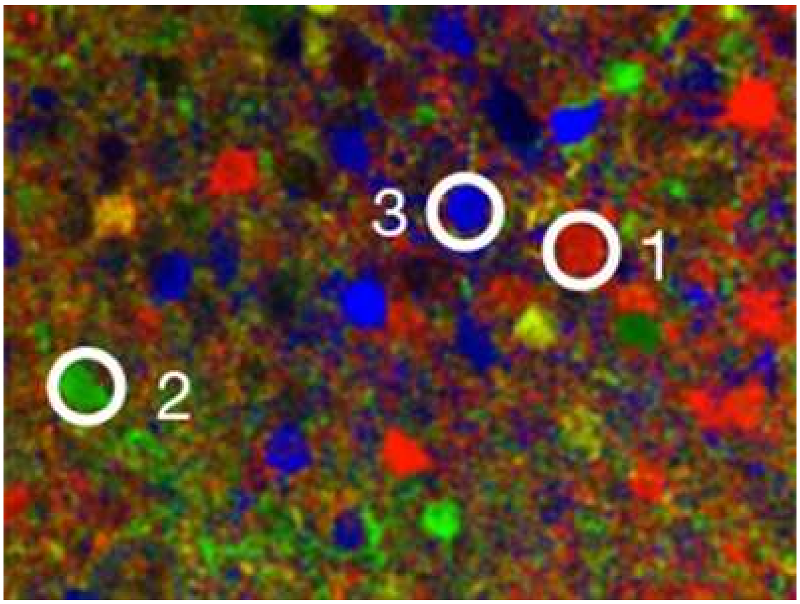


Figure 1 | Structure and response profile of GCaMP8 proteins. a, Zhang *et al.*[1] have developed GCaMP8 — optimized forms of genetically engineered calcium indicators, which respond to intracellular changes in the levels of calcium ions (Ca^{2+}) by fluorescing. GCaMP8 proteins consist of a transmembrane calcium-binding domain (calmodulin), bound to a peptide (a fragment of endothelial nitric oxide synthase; ENOSP) that modifies the activity of calmodulin. Two linkers attach calmodulin to a modified form of green fluorescent protein (GFP). When Ca^{2+} binds, a change in protein configuration leads to GFP fluorescence. **b**, The GCaMP8 sensors (8s and 8f) show several improved qualities compared with existing GCaMP sensors, including higher sensitivity (shown as a change in relative brightness from the protein's baseline) and faster decay kinetics (a key parameter in response rate, shown as an inverse scale in 1/seconds).

Despite the potential of these calcium sensors, early versions of GCaMP were hampered by issues such as poor response kinetics, low calcium sensitivity and limited correlation with the electrical activity of neurons. Initial versions of the sensors could detect only large changes in Ca^{2+} brought about by tens of action potentials, with response kinetics on the order of several hundred milliseconds.

Considerable effort has since been dedicated to improving the performance of GCaMP sensors through structure-guided design. Later versions of GCaMP have even been optimized for sensitivity or speed, with ‘slow’ variants maximizing signal strength, and ‘fast’ variants maximizing response kinetics. Fast variants of GCaMP can now detect changes in calcium in time frames of tens of milliseconds. Other GECIs have also been developed that emit different wavelengths of light in response to Ca^{2+}, enabling several populations of neurons to be imaged simultaneously.



[Read the paper: Fast and sensitive GCaMP calcium indicators for imaging neural populations]

These advances have enabled scientists to capitalize on the versatility, specificity and longevity of GECIs. The fact that they are genetically encoded allows them to be expressed continuously in many cells at once. This has facilitated *in vivo* measurements of large populations of cells in flies, rodents and primates. The cell type in which GECIs are expressed can also be controlled genetically, which has helped researchers to understand the variability of neural circuits. For instance, GECIs have been used to delineate subpopulations of neurons in the brain’s cortex that have distinct calcium dynamics and roles in decision-making[8]. They have also been used to measure Ca^{2+} in non-neuronal cells in the brain. For example, calcium imaging of astrocytes — the most abundant non-neuronal cell type in the

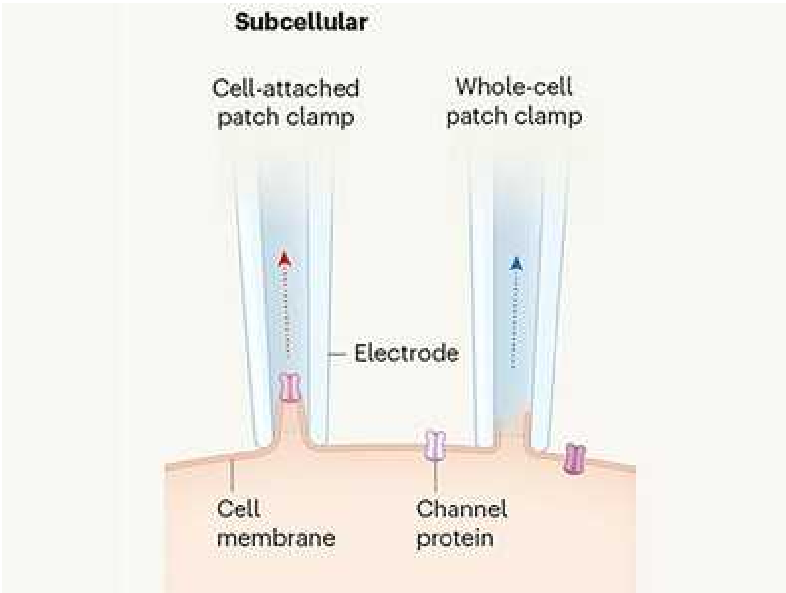
brain — has revealed how neurotransmitters such as dopamine shape the response of astrocytes to neural activity[9,10].

Nonetheless, the latest GCaMP iterations have fallen short of reliably identifying individual action potentials under the 'noisy' conditions in living brains. Several research groups have been racing to identify GCaMP variants that can detect these changes — which occur over just a few milliseconds — without sacrificing brightness. Zhang and colleagues' work introduces a competitor to this race, and points to a bright future for calcium imaging.

YIYANG GONG & CASEY BAKER: Time for the next generation

Single action potentials are the fundamental unit of neural communication. Zhang and colleagues' development is therefore a major advance. The authors' GCaMP8 sensors have significantly faster kinetics than do existing start-of-the-art GECIs (Fig. 1b), such as the GCaMP7 and XCaMP series[11,12]. The GCaMP8 sensors also avoid the typical trade-off between sensitivity and speed, whereby slower sensor kinetics produce larger optical responses.

Zhang and colleagues based their sensors on GCaMP6. The authors optimized this protein by engineering different versions of several of its modules, including the two linkers that connect GFP to the bound peptide and to the calmodulin domain that grabs the peptide when Ca^{2+} binds. They paid particular attention to the interface between the bound peptide (which in GCaMP8 is a fragment of endothelial nitric oxide synthase) and the calmodulin domain, because this region of the sensor has a crucial role in its response and kinetics[12,13]. The team then used an optimization process that involved replacing many amino acids and performing several rounds of testing to identify the best-performing sensors. They validated the performance of these indicators in flies and mice. The strong performance of the GCaMP8 sensors suggests that the way in which Ca^{2+} induces fluorescence in this series is different from that in other GCaMPs.



[A breakthrough method that became vital to neuroscience]

The fast kinetics and high accuracy of the GCaMP8 sensors will enable researchers to analyse phenomena in live animals that could previously have been studied only by using electrical measurements or genetically encoded voltage indicators (GEVIs), which respond to changes in voltage by fluorescing (much like how GECIs respond to calcium). Until now, it has been assumed that GECIs and GEVIs report different types of neural activity — GEVIs determine the timing of action potentials, and GECIs reveal key activity in neuronal compartments (for example, by measuring Ca^{2+} dynamics in processes called dendrites that receive signals from other neurons)[14].

However, with the development of GCaMP8, the information that can be gained from these tools has converged. Zhang and colleagues show, through computational analyses of the light bursts released by GCaMP8, that their sensors can detect action potentials nearly as accurately as can voltage imaging[15,16]. Similarly, GCaMP8 could reliably detect when fly neurons that normally fire frequently become transiently inactive, providing similar results to GEVI-based measurements[17,18].

When both voltage and calcium imaging can be used to record from many neurons simultaneously, practical experimental considerations might persuade future researchers to choose calcium imaging over voltage

imaging. GCaMP8 is compatible with existing microscopy configurations and preparation methods. The series can be used to examine the activity of broad neuronal networks on a larger scale than is possible using voltage imaging. One might imagine that the GCaMP8 series will soon be used to interpret detailed, millisecond-by-millisecond sequences of activity that flow through many neurons in a specific area of the brain.

The fast kinetics associated with this generation of GCaMPs, along with existing GEVIs, will motivate the development of rapid optical-microscopy and imaging technologies. Such microscopes would be able to detect transient bursts of light emitted from many neurons simultaneously. But for now, the fact that calcium-sensor kinetics is no longer the bottleneck in interpreting rapid neural activity will be welcomed by many neuroscientists.

Nature **615**, 804-805 (2023)

doi: https://doi.org/10.1038/d41586-023-00704-y

References

1. Zhang, Y. *et al. Nature* **615**, 884–891 (2023).
2. Shimomura, O., Johnson, F. H. & Saiga, Y. *J. Cell. Comp. Physiol.* **59**, 223–239 (1962).
3. Ridgway, E. B. & Ashley, C. C. *Biochem. Biophys. Res. Commun.* **29**, 229–234 (1967).
4. Llinás, R., Sugimori, M. & Silver, R. B. *Science* **256**, 677–679 (1992).
5. Romoser, V. A., Hinkle, P. M. & Persechini, A. *J. Biol. Chem.* **272**, 13270–13274 (1997).
6. Miyawaki, A. *et al. Nature* **388**, 882–887 (1997).
7. Nakai, J., Ohkura, M. & Imoto, K. *Nature Biotechnol.* **19**, 137–141 (2001).

8. Musall, S. *et al. Nature Neurosci.* **26**, 495–505 (2023).

9. Pittolo, S. *et al. Cell Rep.* **40**, 111426 (2022).

10. Corkrum, M. *et al. Neuron* **105**, 1036–1047 (2020).

11. Dana, H. *et al. Nature Methods* **16**, 649–657 (2019).

12. Inoue, M. *et al. Cell* **177**, 1346–1360 (2019).

13. Inoue, M. *et al. Nature Methods* **12**, 64–70 (2015).

14. Gambino, F. *et al. Nature* **515**, 116–119 (2014).

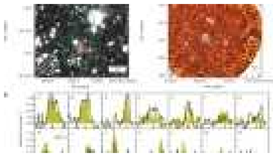
15. Abdelfattah, A. S. *et al. Science* **365**, 699–704 (2019).

16. Villette, V. *et al. Cell* **179**, 1590–1608 (2019).

17. Yang, H. H. *et al. Cell* **166**, 245–257 (2016).

18. Gong, Y. *et al. Science* **350**, 1361–1366 (2015).

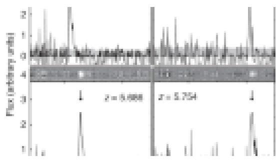
This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00704-y



A massive core for a cluster of galaxies at a redshift of 4.3

25 April 2018

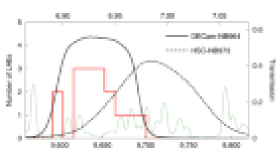
T. B. Miller, S. C. Chapman, … A. L. Strom



A giant protocluster of galaxies at redshift 5.7

15 October 2018

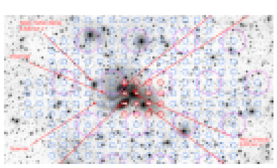
Linhua Jiang, Jin Wu, … Xue-Bing Wu



A Lyman-α protocluster at redshift 6.9

25 January 2021

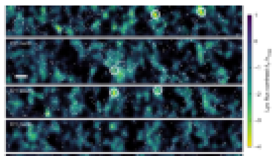
Weida Hu, Junxian Wang, … XianZhong Zheng



A 0.6 Mpc H i structure associated with Stephan’s Quintet

19 October 2022

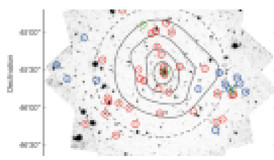
C. K. Xu, C. Cheng, … F. Renaud



[A population of ultraviolet-dim protoclusters detected in absorption]

15 June 2022

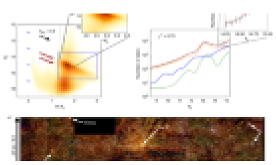
Andrew B. Newman, Gwen C. Rudie, … John S. Mulchaey



[Spectroscopic confirmation of a mature galaxy cluster at a redshift of 2]

01 January 2020

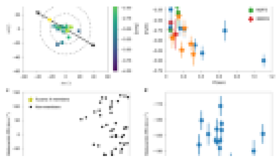
J. P. Willis, R. E. A. Canning, … G. Brammer



[Early formation and recent starburst activity in the nuclear disk of the Milky Way]

16 December 2019

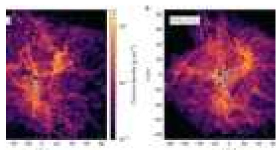
Francisco Nogueras-Lara, Rainer Schödel, … Santi Cassisi



[An extended halo around an ancient dwarf galaxy](#)

01 February 2021

Anirudh Chiti, Anna Frebel, … John E. Norris



[A universal route for the formation of massive star clusters in giant molecular clouds](#)

25 June 2018

Corey S. Howard, Ralph E. Pudritz & William E. Harris

- Article
- Open Access
- Published: 29 March 2023

Forming intracluster gas in a galaxy protocluster at a redshift of 2.16

- Luca Di Mascolo ,
- Alexandro Saro,
- Tony Mroczkowski,
- Stefano Borgani,

- Eugene Churazov,
- Elena Rasia,
- Paolo Tozzi,
- Helmut Dannerbauer,
- Kaustuv Basu,
- Christopher L. Carilli,
- Michele Ginolfi,
- George Miley,
- Mario Nonino,
- Maurilio Pannella,
- Laura Pentericci &
- …
- Francesca Rizzo

Show authors

Nature **volume 615**, pages 809–812 (2023)

- 2043 Accesses

- 220 Altmetric

- Metrics details

Abstract

Galaxy clusters are the most massive gravitationally bound structures in the Universe, comprising thousands of galaxies and pervaded by a diffuse, hot intracluster medium (ICM) that dominates the baryonic content of these systems. The formation and evolution of the ICM across cosmic time[1] is thought to be driven by the continuous accretion of matter from the large-scale filamentary surroundings and energetic merger events with other clusters or groups. Until now, however, direct observations of the intracluster gas have been limited only to mature clusters in the later three-quarters of the history of the Universe, and we have been lacking a direct view of the hot, thermalized cluster atmosphere at the epoch when the first massive clusters formed. Here we report the detection (about 6σ) of the

thermal Sunyaev–Zeldovich (SZ) effect[2] in the direction of a protocluster. In fact, the SZ signal reveals the ICM thermal energy in a way that is insensitive to cosmological dimming, making it ideal for tracing the thermal history of cosmic structures[3]. This result indicates the presence of a nascent ICM within the Spiderweb protocluster at redshift $z = 2.156$, around 10 billion years ago. The amplitude and morphology of the detected signal show that the SZ effect from the protocluster is lower than expected from dynamical considerations and comparable with that of lower-redshift group-scale systems, consistent with expectations for a dynamically active progenitor of a local galaxy cluster.

Main

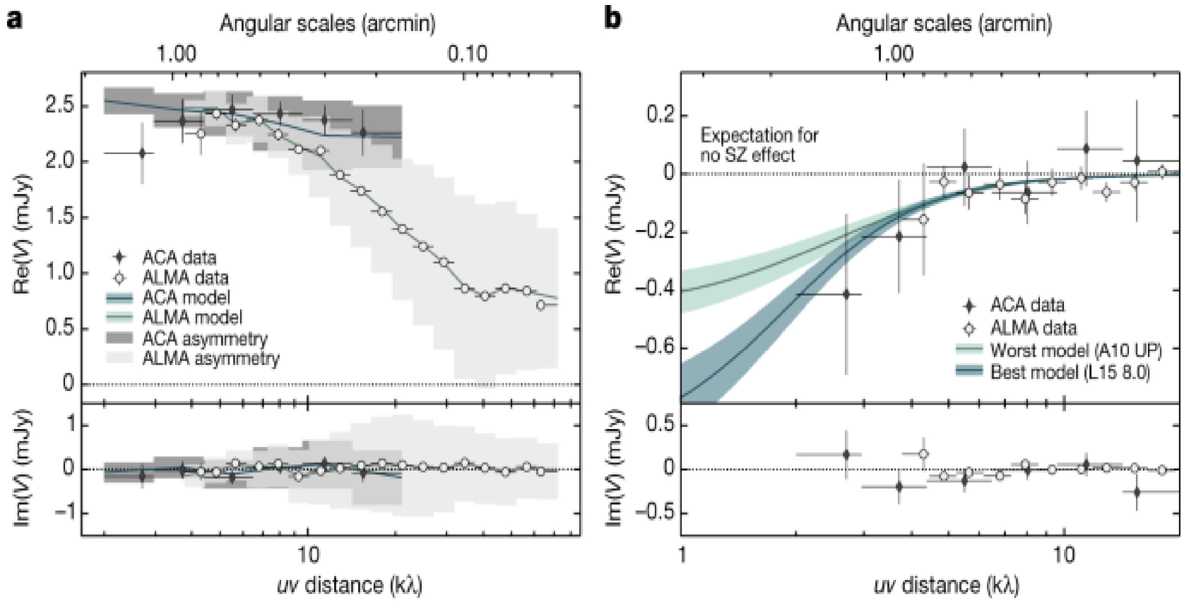
To measure the SZ effect of the protocluster complex surrounding PKS 1138–262 ($z = 2.156$; commonly known as the Spiderweb galaxy), we used the Atacama Large Millimeter/submillimeter Array and obtained deep Band 3 (94.5–110.5 GHz) observations, exploiting both the main 12-m array (ALMA) and the 7-m Atacama Compact Array (ACA).

The transition from sparse protocluster complexes to mature, nearly virialized systems is a tumultuous process[1,4]. Energetic events—for example, infall and accretion of the diffuse medium from surrounding filaments, mergers with substructures and feedback from active galactic nuclei (AGN)—affect the regularity of the assembling ICM, with simulations predicting that their effects could persist for more than a Hubble time[5]. From an observational point of view, this implies that building a simple, analytical model describing the morphology of the disturbed proto-ICM is not a trivial task. Further, the possibility of mapping the ICM within galaxy protoclusters relies on our ability to separate the SZ footprint from any contaminating sources within the field. In fact, the Spiderweb galaxy harbours a powerful AGN[6,7], with associated hybrid-morphology[8,9] jets extending over scales of about 100 kpc. The result is that the millimetre-wavelength continuum signal in the direction of the Spiderweb protocluster is dominated by the emission from the central radio galaxy. In the available Band 3 ACA data (probing the SZ signal thanks to its ability to recover larger angular scales), we measure a peak surface brightness for the continuum emission from the central radio source of 2.39 ± 0.12 mJy

beam^{-1}. By assuming that, also at the Spiderweb redshift, the total mass and volume-integrated SZ signal scale covariantly (as established theoretically and empirically for low-z clusters; see ref. [10] for a review), we find that such a flux estimate is at least an order of magnitude larger than the absolute amplitude of the peak SZ signal expected for the Spiderweb protocluster ($\lesssim$0.2 mJy beam^{-1}) computed by considering an upper limit of about 10^{14} $M_{\odot}$ for the mass of the system[11,12,13,14,15] (with $M_{\odot}$ denoting the mass of the Sun). This large difference in the dynamic range of the surface brightness of the extended structure of the Spiderweb galaxy and of the underlying SZ effect limits the possibility of performing a robust separation of the two signals through standard imaging techniques. Thus, to handle the above complexities and obtain a statistically robust detection of the SZ signal, we need to rely on simplifying assumptions. We thus assume that the SZ signal is generated by a spherically symmetric ICM distribution and that the extended radio source can be described by a collection of point-like components (Methods). We then analysed the available ALMA data using a Bayesian forward modelling approach[16]. The inclusion of a model component for the SZ signal is favoured by the Bayesian evidence (that is, the normalization factor in Bayes' theorem, given by the likelihood marginalized over the prior volume and key element for performing Bayesian model selection) over the case comprising only the jet and AGN emission (Fig. 1) at an effective significance $\sigma_{\rm eff} = 5.97 \pm 0.08$. This is estimated from the difference $\Delta \log \,Z$ of the logarithm of the evidences (hereafter, log-evidence) for the models with and without a SZ component and by assuming the posterior distribution to be described by a multivariate normal distribution (that is, ${{\sigma }}_{{\rm{e}}{\rm{f}}{\rm{f}}}={\rm{s}}{\rm{g}}{\rm{n}}(\Delta \log \,Z)\cdot \sqrt{2|\Delta \log \,Z|}$). We further note that the reported value represents a conservative lower limit on the actual significance of the SZ signal, as different prescriptions for the pressure distribution of the ICM further improve the Bayesian evidence (see Extended Data Table 2 in Methods). Given the above self-similar assumption for the pressure distribution, we estimate that the detected SZ signal would correspond to a halo with ${M}_{500}=(3.4{6}_{-0.43}^{+0.38})\times 1{0}^{13}\,{M}_{\odot }$ and ${r}_{500}=228\,.{\,9}_{-9.5}^{+8.4}\,{\rm{kpc}}$, in which M_{500} is the total mass contained in the spherical volume of radius r_{500} within which

the average density is 500 times larger than the critical density of the Universe at the source redshift (here, as in the rest of the manuscript, the best-fit value and the uncertainties correspond to the 50th, and 16th and 84th percentiles of the posterior distribution function for a given model parameter, respectively). At face value, this mass constraint is much lower than the naive (and largely uncertain) expectations based on velocity-dispersion measurements previously reported in the literature[11,14,15]. We find that considering different assumptions for the ICM pressure distributions slightly relieves this tension but produce mass estimates still smaller than expected from dynamical considerations. However, comparing our estimate of the volume-integrated SZ signal with the corresponding value predicted from dynamical mass estimates provides an empirical demonstration that the detected ICM halo is probably part of an extended complex of several interacting substructures. Similar hints are observed when repeating our analysis by including several SZ components to the overall model, without however providing conclusive and statistically meaningful results. The observed discrepancy between the SZ mass estimate and the value from velocity-dispersion measurements, in combination with the evidence for a positional offset between the Spiderweb galaxy and the SZ centroid is thus consistent with a merging scenario, in which the complex structure of the pressure distribution of the proto-ICM is not necessarily well captured by the simple analytical models used in our analysis. A thorough discussion of the potential explanations of such discrepancy, together with the assessment of potential systematic issues, can be found in Methods.

Fig. 1: Binned *uv* profiles of the Band 3 ALMA and ACA data.



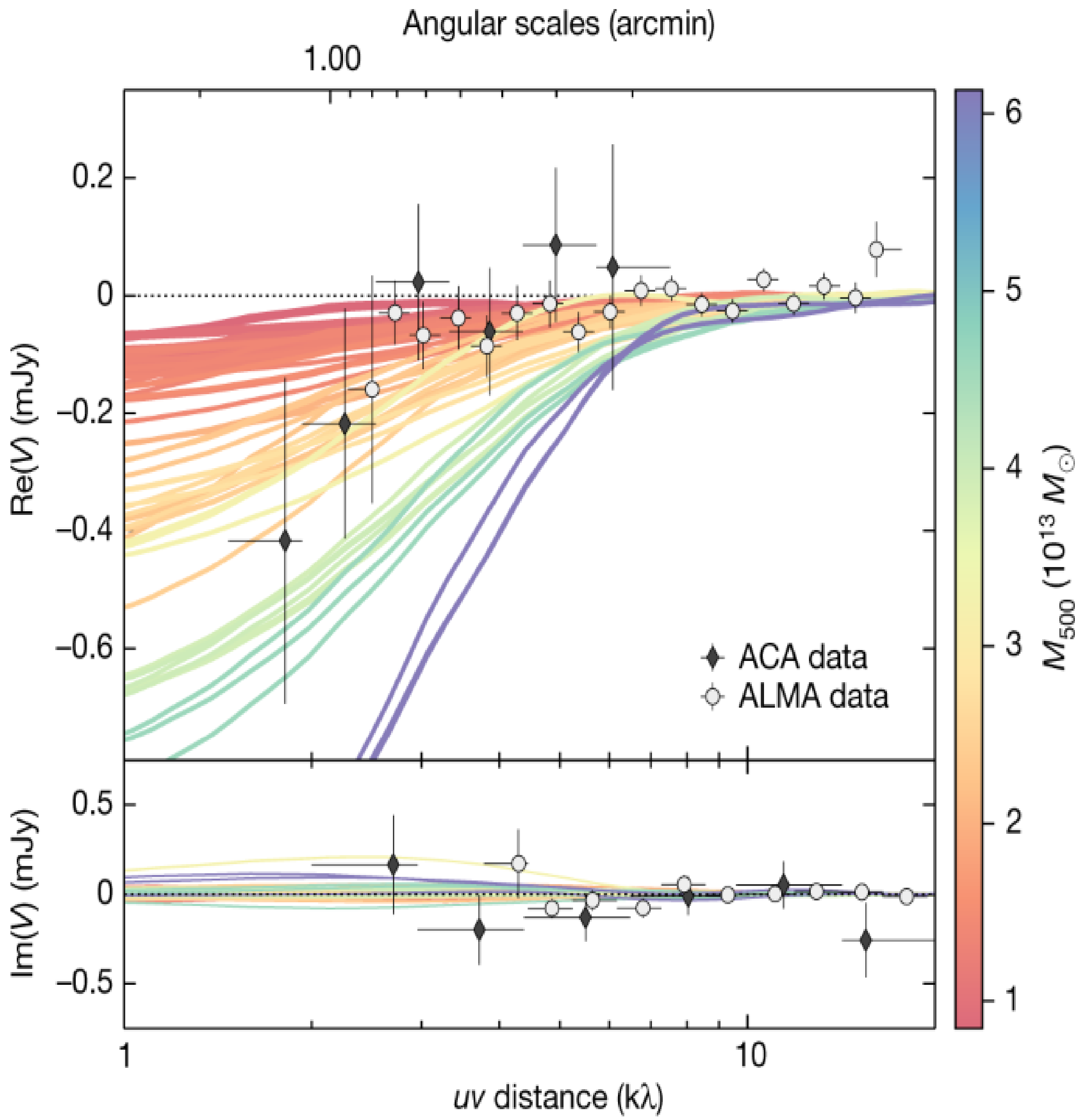
a, Comparison of the real (Re) and imaginary (Im) parts of the visibilities (*V*) for the binned ALMA and ACA data, and for the radio-source model. The solid lines represent the corresponding median *uv* profile for each visibility set obtained by marginalizing over different numbers of point-like components (see Methods). The model uncertainties are computed accordingly but, despite being plotted, are too small to be visible. The shaded regions denote the standard deviation of the azimuthal variation of the model amplitude in each *uv* bin, owing to the elongated morphology of the radio galaxy and the asymmetric visibility patterns. The systematic drop in the real component of the visibilities at short *uv* distances (that is, large angular scales) provides evidence for the presence of a SZ decrement towards the Spiderweb complex. We note that, as we do not see a similar deviation in the imaginary component, this cannot be ascribed merely to phase variations owing to off-centre sources. **b**. Comparison of SZ models with the ALMA and ACA data. Shown are the worst (A10 UP; ref. [28]) and best (L15 8.0; ref. [29]) SZ profiles, intended as the models that are statistically favoured the least and the most, respectively, based on their Bayesian evidence. Before binning the data, we subtracted the median radio model from the visibilities and shifted the phase centre on the SZ centroid. The uncertainties associated with the radio model and with the SZ coordinates are propagated into the ones for the ALMA and ACA data points. The shaded regions correspond to the 68% credible intervals for each model. Also in this case, the uncertainties and median SZ profiles are

marginalized over models with varying numbers of point-like components. The divergence of the two models and the increase of their uncertainties at small *uv* distances (that is, large scales) is symptomatic of the limited capabilities of ALMA+ACA in constraining fluxes above about 1.70′. For both panels, the error bars denote 1σ uncertainties.

Source data

To circumvent the limitations of these analytic results and corroborate our measurements with a more realistic and physically complex model, we perform a comparison of the SZ signal that we reconstructed from the ALMA+ACA data with mock SZ observations based on cosmological hydrodynamical simulations[17,18] of galaxy protoclusters (Methods). We emphasize that the scope of this comparison is not aimed at identifying an exact simulation counterpart to the Spiderweb protocluster but rather at guiding our interpretation of the measured SZ effect by providing quantitative predictions on the overall SZ signal expected to be measurable in the observations. We generate mock ALMA+ACA observations for a set of 27 simulated massive halos ($M_{500} \gtrsim 1.3 \times 10^{13}\, M_{\odot}$ at $z = 2.16$) that represent progenitors of galaxy clusters with masses at redshift $z = 0$ in the range $M_{500} = (5.6–8.8) \times 10^{14}\, M_{\odot}\, \mathrm{h}^{-1}$. In Fig. 2, the resulting *uv* profiles for all the simulated clusters are compared with the radio-source-subtracted ALMA+ACA data (as shown in Fig. 1). These show an agreement between the observed SZ signal, the prediction for halos of mass $M_{500} \simeq (2–5) \times 10^{13}\, M_{\odot}$ and, in turn, the independent estimates of the protocluster mass from the parametric modelling. Such a result provides a straightforward assessment of the reliability of the above results and, therefore, of the mismatch of the amplitude of the observed SZ signal with the expectations from dynamical considerations.

Fig. 2: *uv* profiles from mock SZ observations of simulated galaxy protoclusters.



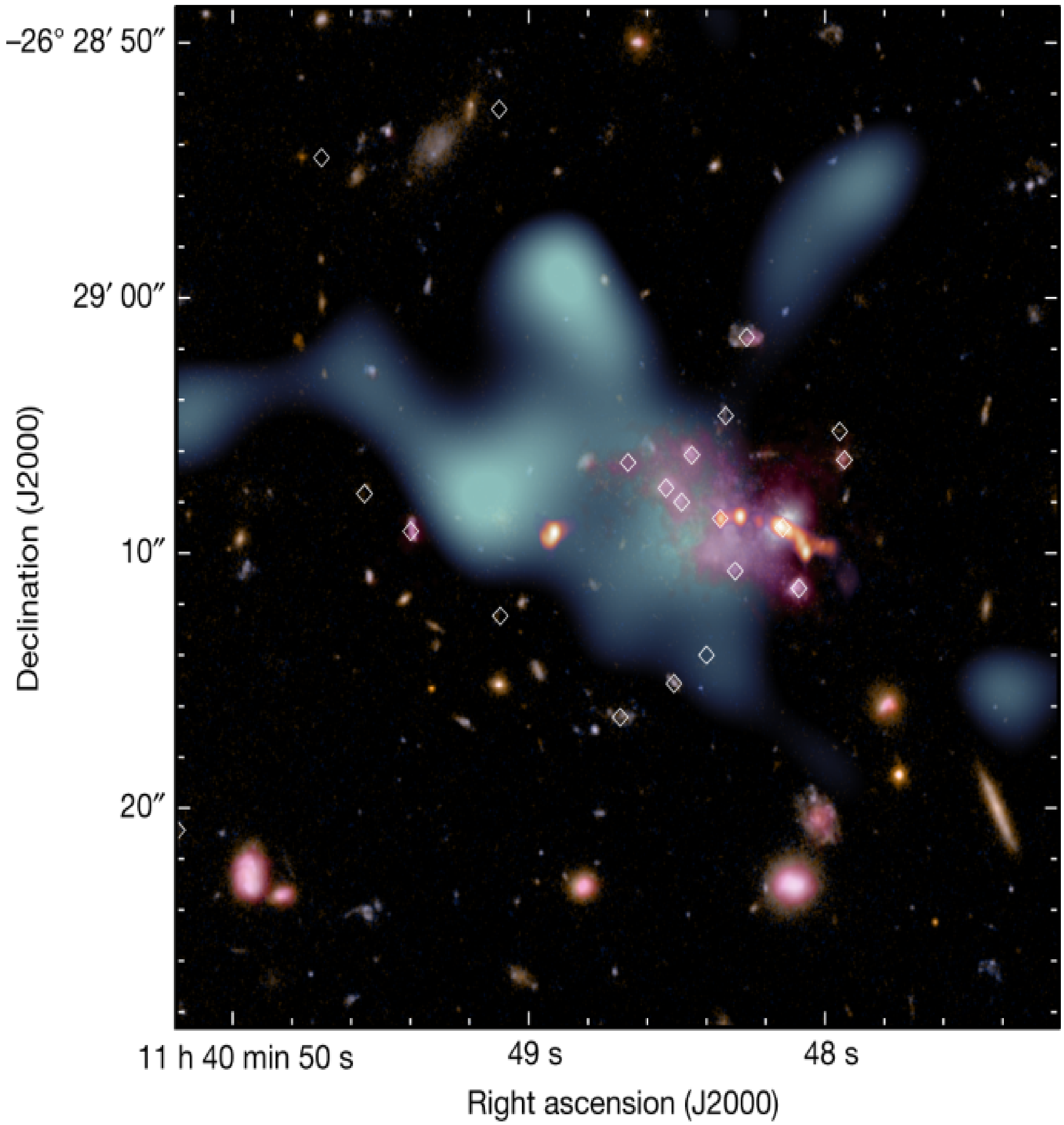
Each curve represents the *uv* profile obtained for a halo extracted from the Dianoga set after projecting the corresponding SZ map onto the visibility plane, colour-coded according to their mass. For a comparison, we further plot the binned ACA and ALMA measurements from Extended Data Fig. 6. Consistent with our finding from the parametric modelling, the observed SZ signal suggests that the Spiderweb protocluster is characterized by a mass $M_{500} \simeq (2\text{–}5) \times 10^{13}\ M_{\odot}$. We note that this result is only weakly dependent on the cosmological model adopted in the reference simulation. As for Fig. 1, the error bars correspond to the 1σ uncertainties on the binned ALMA and

ACA data. Re(V) and Im(V) are the real and imaginary parts of the visibilities (V), respectively, for the binned ALMA and ACA data.

Source data

The analysis described above required a reduction of the computational complexity by limiting the range of *uv* scales used for the model reconstruction. To obtain a high-resolution view of the Spiderweb complex and fully exploit the entirety of the dynamic range of physical scales investigated by ALMA+ACA, we apply an independent, sparse modelling approach to image the available data (see Methods for details). The resulting high-resolution image of the SZ signal obtained after subtracting the best-fit radio source from the ALMA+ACA data is provided in Fig. 3, along with a multiwavelength perspective on the Spiderweb complex (see also Extended Data Fig. 6 in Methods). Overall, the inclusion of this information in the context of the wealth of multiwavelength data associated with the Spiderweb system strongly supports an extremely dynamically active phase of protocluster formation (Methods). The multiphase environment is in fact experiencing complex interactions with the extended radio galaxy and mass accretion from the large-scale cosmic web and energetic merging events[15,19], consistent with the considerations from our parametric analysis.

Fig. 3: Multiwavelength view of the Spiderweb complex.



Composite Hubble Space Telescope image based on ACS/WFC F475W and F814W data of the Spiderweb field. Overlaid are the emission from the Spiderweb galaxy and associated extended radio jet as measured by the VLA in X-band[8,9] (8–10 GHz; red), the image of the extended Ly*α* nebula[19,30] observed with the FORS1 instrument on the VLT (pink) and the SZ signal from a combined ALMA+ACA image (light blue). When imaging the ALMA+ACA data, we applied to the visibility weights a *uv* taper with $\sigma_{taper} = 20\ k\lambda$ to both suppress any noise structures on small scales and to emphasize the bulk distribution of the SZ signal from the ICM of the

Spiderweb protocluster. The SZ effect offset with respect to the Spiderweb galaxy suggests that the protocluster core is undergoing a dynamically active phase. The white diamonds denote all the spectroscopically confirmed member galaxies summarized in ref. [25] and references therein.

Overall, the reported identification in the ALMA+ACA observations of the SZ signal from the central region of the Spiderweb complex is providing the direct indication that the system is more than a loose association of galaxies and has already started assembling its own halo of diffuse intracluster gas. Most importantly, this is providing a statistically meaningful confirmation of long-standing predictions from cosmological simulations[13] for the existence of an extended halo of thermalizing ICM within the Spiderweb protocluster, as well as of observational works, so far limited just to indirect[8,20,21] evidence or tentative detections[7,22,23]. At the same time, this detection shows that current SZ facilities could be used to effectively open a new observational window on protocluster environments. Many such systems, including more massive protoclusters, are in fact expected[4] to exist out to $z \simeq 3$. To this end, the SZ effect provides the means for obtaining a straightforward, unambiguous confirmation of the probable progenitors of local galaxy groups and clusters, as well as of the continuing thermalization of the forming ICM in these systems or any of their parts[24]. In fact, even in the case of the extensively studied Spiderweb protocluster—for which we have spectroscopic confirmation for 112 member galaxies across the entire protocluster structure[25]—a robust identification of distinct subhalos through a spectroscopic characterization has so far been ineffective. Similarly, obtaining a robust identification of a thermal ICM component with current X-ray facilities would require a prohibitive amount of observing time, as well as an accurate separation of thermal, inverse Compton and AGN contributions to the overall X-ray signal[22,23,26,27]. We refer, for instance, to the recent Chandra study[23] of the Spiderweb protocluster (see also Extended Data Fig. 6 in Methods), which, despite being based on observations with an exposure time more than an order of magnitude larger than the ALMA+ACA measurements used in this work, provided tentative but inconclusive support to the thermal origin of a diffuse component in the X-ray emission. Identifying and performing detailed characterization of many more of the first-forming clusters during the crucial phase of vigorous relaxation and thermalization will be essential for gaining a comprehensive

view of the emergence of galaxy clusters from the large-scale structure of the Universe, as well as the environmental processing of galaxies in the earliest phases of their evolution. Similarly, it will help shed light on the role of feedback mechanisms in determining the physical properties of the diffuse baryons in the proto-ICM at the very epochs when their activity is expected to peak and to have a key impact on galaxy and protocluster formation.

Methods

Cosmology

In this work, we consider a spatially flat Λ Cold Dark Matter cosmological model, with $\Omega_M = 0.30$, $\Omega_\Lambda = 0.70$ and $H_0 = 70.0$ km s^{-1} Mpc^{-1}. At the redshift of the Spiderweb complex ($z = 2.156$), 1″ corresponds to 8.29 kpc.

ALMA observations and reduction

An extensive observational campaign was performed during Cycle 6 to obtain a detailed view of the Spiderweb complex in Band 3 (project code: 2018.1.01526.S, principal investigator A. Saro). The data comprise measurements from the main 12-m ALMA[31] array in three different configurations (C43-1, C43-3 and C43-6), aimed at providing a high-dynamic-range view of the structure, as well as from the 7-m ACA[32] (also known as the Morita Array), complementing the ALMA observations over SZ-relevant scales (that is, over a *uv* range of 2.2–17.5 kλ, corresponding to scales 767–97 kpc at the redshift of the Spiderweb galaxy). The spectral setup for all the configurations was tuned to cover the frequency range 94.5–110.5 GHz, split over four 2-GHz-wide spectral bands centred at 95.5 GHz, 97.5 GHz, 107.5 GHz and 109.5 GHz, respectively. In particular, the last window targets the line emission resulting from the $J = 3–2$ transition from the carbon monoxide (CO; rest frequency 345.796 GHz). As we are interested in modelling only the continuum component of the observed signal, we conservatively exclude all the visibilities from the spectral window expected to contain the redshifted CO $J = 3–2$ line. In fact, excluding only the channels corresponding to the specific emission line would provide a slight improvement in the overall statistics of the analysed data. Previous studies[33,34,35] of the molecular content of the nuclear region around the Spiderweb galaxy and of protocluster galaxies have shown, however, that the cold molecular gas within the Spiderweb complex is characterized by large velocity dispersion as well as broad differences in the systemic velocities of the member galaxies. Faint tails and offset components may hence contaminate the continuum signal in any channels in the proximity of the emission line, potentially affecting the model reconstruction.

Data calibration was performed in the Common Astronomy Software Applications[36] (CASA; https://casa.nrao.edu/) package version 5.4.0 using the standard reduction pipeline provided as part of the data delivery. The direct inspection of output visibility tables highlighted no clear issues with the outcome of the pipeline calibration and we therefore did not perform any extra flagging or postprocessing tuning. The resulting root-mean-square

(RMS) noise levels of the observations are estimated to be 4.10 μJy beam^{-1}, 29.1 μJy beam^{-1} and 14.2 μJy beam^{-1} for the C43-1, C43-3 and C43-6 ALMA measurements, respectively, and 32.0 μJy beam^{-1} for the ACA data.

To obtain better knowledge of the spectral properties of the measured signals, we further include in our analyses archival Band 4 ALMA (project code: 2015.1.00851.S, principal investigator B. Emonts) and ACA (project code 2016.2.00048.S, principal investigator B. Emonts) observations. In both cases, we use the calibrated measurement sets provided by the European ALMA Regional Centre network[37] through the calMS service[38]. The achieved RMS noise levels amount to 99.0 μJy beam^{-1} and 6.29 μJy beam^{-1} for the ACA and ALMA data, respectively. As with the Band 3 measurements, we exclude from our analyses the spectral windows in Band 4 covering the CO $J = 4–3$ and [CI] $^3P_1–^3P_0$ emission lines[33].

All interferometric images presented in this work are generated using CASA package version 6.3.0.

Nested posterior sampling

To obtain a statistically robust detection of the potential SZ signal in the direction of the Spiderweb complex, we use the approach already used in refs. [16,39,40,41] in the context of ALMA+ACA studies of the SZ signal. In brief, we perform a visibility-space analysis, which allows for exactly accounting for the non-uniform radio-interferometric transfer function, as well as taking advantage of the Gaussian properties of noise in the native Fourier space. Any extended model component is created in image space, taking into account the proper frequency scaling and primary-beam attenuation for any fields and spectral windows used in the analysis, and is then projected onto the visibility points by means of a non-uniform fast Fourier transform algorithm based on convolutional gridding (as implemented in the finufft library[42]). Instead, in Fourier space, point-like sources are trivially represented by a constant function with amplitude equal to the source flux corrected for the primary-beam attenuation at the source position and with a phase term defined by the offset between the point source and the phase centre of the observations. To allow for Bayesian

model selection and averaging, the sampling of the posterior distribution is performed by means of the nested sampling algorithm[43,44]. We specifically exploit the implementation provided in the dynesty (ref. [45]) package, which allows for robustly extending the sampling problem to moderate-dimensional and high-dimensional models. For any further details on the model reconstruction, we refer to the discussion provided by Di Mascolo et al.[16,39].

Obtaining a thorough description of the small-scale, complex morphology of the extended radio signal from the Spiderweb galaxy would require performing a pixel-level model inference (see ‘Sparse imaging’ section below), resulting in a posterior probability function with extreme dimensionality. Nested sampling techniques generally show better performances than Monte Carlo Markov chain algorithms in the case of moderate-dimensional problems (mitigating the impacts of the so-called ‘curse of dimensionality[44,46]). Still, sampling from high-dimensional posterior distributions may easily become computationally intractable owing to the complexity of estimating high-dimensional marginal likelihoods. Therefore, we decide to perform a first modelling run only on a large-scale subset of the available data. The giant Lyα nebula observed to surround the Spiderweb galaxy is in fact expected to be confined within a diffuse halo of hot intracluster gas[8,9,47,48,49]. In turn, the SZ footprint of potential intracluster gas within the Spiderweb protocluster should be expected to extend over characteristic scales $\gtrsim 10''$. We therefore introduce an upper cut in the visibility space at a *uv* distance of 65 kλ, whose corresponding angular scale is roughly twice the transverse width of the jet structure (that is, the size measured along the direction perpendicular to the jet direction) observed when imaging only the high-resolution ALMA data from the C43-3 and C43-6 observations. Such a choice makes the jet signal spatially resolved only along the jet axis and allows for describing this as a limited collection of point-like sources. As a result, the number of parameters required to model the observations remains limited and the analysis computationally manageable, while allowing for using nested sampling to obtain statistically meaningful information on our model inference.

The selection of the specific number of compact components required to model the radio source was performed by means of Bayesian model selection. In particular, we consider as the most favoured model set the one for which the introduction of an extra point-like term would have caused a degradation or only a marginal improvement in the log-evidence[50] (that is, $(\log \,{Z}_{n+1}-\log \,{Z}_{n})\lesssim 0.50$, in which *n* denotes the number of model components). For all of the components, as already described above, we assume the spatial morphology to be described by a Dirac δ function (that is, a constant function in Fourier space with non-null phase term) and the source fluxes by a power-law spectral scaling. The priors for all the parameters—right ascension, declination, flux and spectral index—are described by uniform probability distributions. In particular, the right ascension and declination are allowed to vary within the area of $r \simeq 1'$ enclosed within the first null of the antenna pattern for the Band 3 ALMA data at the highest available frequency (that is, 107.5 GHz). To avoid label switching and force mode identifiability, we impose a further ordering prescription[51] to the right ascension parameters. The flux and spectral index parameters are instead allowed to vary within uniform prior probability distributions. In particular, we constrain the source fluxes to be non-negative and assume an upper limit of 10 mJy on each amplitude, around an order of magnitude larger than the emission peak in the ALMA map (1.54 mJy; see top panel of Extended Data Fig. 1). For the spectral indices, we consider a range [−5, 10], arbitrarily wide and set to cover both the cases of power spectra with negative and positive slopes, consistent with synchrotron-like and dust-like spectral properties, respectively. The prior limits are intentionally set to extend well beyond the values expected for such cases, to avoid overconstraining of the source spectral properties while allowing for a quick diagnostics of the actual constraining power of the available data with respect to spectral information.

To describe the pressure distribution of the ICM, we instead use a generalized Navarro–Frenk–White (gNFW) profile[52], widely shown to provide an accurate description of the average pressure distribution of the intracluster gas. In particular, we use different gNFW formulations from the literature. A summary is provided below.

- –The universal profile (hereafter, A10 UP) by Arnaud et al.[28] derived from the reconstruction of the pressure distribution in galaxy clusters from the REXCESS[53] sample and the equivalent model obtained from the subset of systems with clear evidence of a disturbed morphology (A10 MD). Although this profile is calibrated on massive local systems ($10^{14}\,M_{\odot} < M_{500} < 10^{15}\,M_{\odot}$, $z < 0.2$), it is the base model used for the mass reconstruction in large-scale SZ surveys and allows for a straightforward comparison with the literature.

- –The pressure model[54] (M14 UP) obtained from the X-ray analysis of a high-redshift ($0.6 < z < 1.2$) subsample of galaxy clusters detected by the South Pole Telescope[55,56,57] (SPT). As for the previous case, we further consider the pressure profile (M14 NCC) reconstructed by excluding all the systems with clear evidence for the presence of a cool core, generally indicative of a more relaxed dynamical state. To our knowledge, this model represents the highest-redshift, observationally motivated pressure profile available at present, in turn potentially providing a better description of the pressure distribution in the Spiderweb system than the A10 models.

- –The median pressure profiles reconstructed from the OverWhelmingly Large Simulations (OWLS) suite of cosmological hydrodynamical simulations, cosmo-OWLS[29], and considering different prescriptions for the physical model of AGN-driven heating of the intracluster gas. In particular, we consider the OWLS[58] reference model (L15 REF), as well as the two AGN models (L15 8.0 and L15 8.5; we refer to Le Brun et al.[29] for details). These models present two main advantages. First, they are built on simulated halos whose mass and redshift ranges ($2 \times 10^{12}\,M_{\odot} \lesssim M_{500} \lesssim 3 \times 10^{15}\,M_{\odot}$, $0 < z < 3$) broadly overlap with the properties of the Spiderweb protocluster. Second, at $z \simeq 2$, the Spiderweb complex sits in a transitional phase for AGN feedback[59,60] and the different flavours of the cosmo-OWLS pressure model allow for directly using the SZ effect to explore different AGN scenarios.

- –The mass-dependent and redshift-dependent extended pressure model[61] (G17 EXT) calibrated on simulated galaxy clusters from the set of Magneticum Pathfinder hydrodynamical simulations (http://www.magneticum.org/). Although this profile is computed on massive galaxy clusters ($M_{500} > 1.4 \times 10^{14}\,M_{\odot}$), it provides an explicit model for taking into account the departure from universality and self-similarity as a function of mass and redshift.

In all cases, the free parameters defining the gNFW profiles are the plane-of-sky coordinates of the model centroid and the mass parameter M_{500}. As for the radio model, the right ascension and declination parameters are prescribed to vary within the region encompassed by the first null of the Band 3 ALMA primary beam. For the mass M_{500}, we consider a log-uniform distribution over the range $[10^{12}\,M_{\odot}, 10^{15}\,M_{\odot}]$, to facilitate the posterior exploration over such a wide range of order of magnitudes.

We also tried fitting the cool-core versions of the A10 and M14 profiles above but found these to be systematically disfavoured (\(\Delta \log \,Z\lesssim 15.5\)) in comparison with the listed models. This is, however, not surprising, as the presence of a well-formed cool core would be hardly consistent with the inherently disturbed nature of a protocluster complex.

Finally, we account for any potential systematics with data calibration by considering a scaling parameter for each of the measurement sets used in our analysis. For these, we assume normal prior distributions, with unitary central value and standard deviation of 5%, as reported in the ALMA Technical Handbook for the considered observing cycles.

Results

With regards to the extended signal from the Spiderweb galaxy, the criterion introduced above for the selection of the number of point-like components supports the case for a total of eight distinct components over the entire search area (we summarize the key information on the results of the point-like modelling in Extended Data Table 1). In particular, two components (ID1 and ID2) are found to be spatially consistent with the

position of known protocluster members[25,34,35,49,62,63,64]—that is, ERO 284 (ref. [12]) and HAE 229 (refs. [65,66]), respectively—located around 250 kpc west of the Spiderweb galaxy. The remaining components instead uniquely describe the radio emission associated with the Spiderweb galaxy, with one point-like component (ID8) specifically corresponding to the bright lobe of the eastern radio jet and another (ID6) being nearly coincident with the Spiderweb galaxy itself (Extended Data Fig. 1). All the components describing the extended signal exhibit a negative spectral index, consistent with the synchrotron origin of the emission. The best-fit estimates highlight a spatial variation consistent with what is observed at lower frequencies in the Karl G. Jansky Very Large Array (VLA) data[6,7,8,9], which show that the spectrum of the eastern lobe is, on average, steeper than that of the Spiderweb galaxy and the western jet. The specific values are also in rough agreement with the results from the VLA analyses. A one-to-one comparison is, however, not practicable, owing to the inherent high-frequency spectral steepening induced by radiative losses and the different modelling approaches. The two offset sources, on the other hand, are both characterized by positive spectral indices. Such a trend is possibly indicating a dominant contribution from thermal dust emission already at about 100 GHz and is in agreement with the potential presence of massive dust reservoirs in the galaxies, as already verified for HAE 229 (refs. [65,66]).

A summary of the inferred parameters for the different SZ models is instead provided in Extended Data Table 2. Despite the more or less substantial differences in the reconstructed parameters, all the adopted pressure profiles resulted in statistically consistent SZ models, not allowing a statistically motivated selection of a specific description. All the assumed models, with exception of the L15 8.0 and L15 8.5 cases, provide mass estimates $M_{500} \simeq 3 \times 10^{13}\,M_{\odot}$. In fact, we note that the L15 8.5 profile from Le Brun et al.[29] results in a mass of about $8 \times 10^{13}\,M_{\odot}$, consistent with many of the dynamical estimates reported in the literature[11,13,15] for the Spiderweb protocluster. Considering that the profile is based on simulations with the prescription for an intense heating of the ICM owing to AGN feedback, the result might hint to a crucial role of the active core of the Spiderweb galaxy in shaping the intracluster/circumgalactic medium. Nevertheless, this model is statistically equivalent to many others in our sample, limiting the

statistical relevance of the above considerations. For the same reason, any attempts of performing a comparison with past ICM studies of similarly high-redshift systems—for example, XLSSC 122 (ref. [67]) or Cl J1449+0856 (ref. [68]), the clusters at the highest redshift known so far with a direct SZ measurement—would not be statistically meaningful.

Finally, we note that, in our analyses, all the scaling parameters are found to be broadly consistent with unity. In particular, the scaling factors are measured to be equal to $1.0{4}_{-0.04}^{+0.04}$ and $1.0{1}_{-0.04}^{+0.03}$ for the Band 3 ACA and ALMA observations, respectively, whereas—in the case of the Band 4 data—we find parameters of $1.0{0}_{-0.07}^{+0.06}$ for ACA and $1.0{3}_{-0.04}^{+0.04}$ for ALMA (the reported values are given by the Bayesian model averages for all models considered in this work; see the ‘Dependence of the SZ significance on the number of point-like components’ section below for a discussion).

Comparison with masses from previous studies

Performing a proper comparison of the SZ-derived mass estimates (Extended Data Table 2) with independent measurements from the literature is non-trivial. The dynamical masses for the Spiderweb protocluster are in fact based on velocity-dispersion estimates that might trace specific, yet not well-identified substructures and that span almost an order of magnitude: from 204 km s^{-1} for one peak in the velocity distribution of Lyα emitters in the Spiderweb field[11] and up to 1,360 km s^{-1} as measured for all satellites within 60 kpc from the Spiderweb galaxy[14]. Nevertheless, the SZ-based mass estimates presented in this work are much lower (a factor of about 2–4, depending on the model) than the dynamical values[11,14,15] found in the literature for the mass of the whole protocluster structure. In fact, the same studies reported evidence for a complex velocity structure within the central region of the Spiderweb complex, hinting at the possibility for the system to be experiencing a major merger and still accreting large amounts of material from surrounding filaments. The fact that the Spiderweb system is embedded in a large-scale filamentary structure was confirmed by wide-field CO $J = 1$–0 mapping[35]. As such, even the very core of the Spiderweb

protocluster might not be fully virialized. Accordingly, the integrated SZ signal $Y_{SZ} = Y_{SZ}\ (<5r_{500})$ we measure from the ALMA+ACA observations —$Y_{\rm SZ}^{\rm ALMA} = (1.68_{-0.32}^{+0.35})\times 10^{-6}\,{\rm Mpc}^{2}$ (see Extended Data Table 2)—is, for instance, a factor of 3.5 times lower than that expected, taking the mass inferred from the velocity-dispersion measurement $\sigma_{\rm v} \simeq 683\ {\rm km\ s^{-1}}$ reported by Shimakawa et al.[15] for the galaxies within the 0.53-Mpc region surrounding the Spiderweb galaxy.

This large difference between the expected SZ signal from velocity-dispersion measurements and the observed ALMA+ACA SZ integrated flux could be the result of a scenario in which the measured SZ signal is dominated by the contribution from the most prominent (sub)halo. In fact, the integrated SZ flux Y_{SZ} scales steeply as a function of mass M, meaning that the SZ signal from a single halo would be larger than the sum of the SZ flux from a complex system of subhalos whose masses amount overall to the same value M. Under the assumptions that the Spiderweb protocluster is composed of several interacting substructures and the measured line-of-sight velocity dispersion $\sigma_{\rm v}$ is providing an unbiased estimate of the total mass of the system, we can exploit the Y_{SZ}–M relation to obtain an estimate of the number $n_{\rm halos}$ of subhalos populating the Spiderweb complex.

First, the dispersion $\sigma_{\rm v}$ is converted into a dynamical mass estimate using the scaling relation calibrated by Saro et al.[69] on a mock galaxy catalogue from the Millennium Simulation[70]. This is rescaled to M_{500} assuming the conversion relation in Ragagnin et al.[71] between masses at different overdensities. We then iterate over $n_{\rm halos}$ and compute the integrated SZ flux expected for each set of subhalos. For the sake of simplicity, we consider all the $n_{\rm halos}$ subcomponents to have equal mass $M_{500}^{\sigma_{\rm v}}\,/n_{\rm halos}$. Both the masses from our SZ analysis $M_{500}^{\rm ALMA}$ and from the velocity-dispersion measurement $M_{500}^{\sigma_{\rm v}}\,/n_{\rm halos}$ are then used to obtain a measurement of the spherically integrated SZ flux Y_{SZ} by means of numerical integration of the pressure profiles over a volume of radius $5r_{500}$.

In our calculations, we consider the universal formulation by Arnaud et al.[28] to describe the pressure distribution of the intracluster electrons. Finally, to derive the number of equal-mass subhalos within the Spiderweb protocluster whose individual SZ fluxes would match the measured value $({Y}_{{\rm{SZ}}}^{{\rm{ALMA}}}={Y}_{{\rm{SZ}}}\left({M}_{500}^{{\rm{ALMA}}}\right))$, we simply estimate the value $n_{\rm halos}$ for which the equality $({Y}_{{\rm{SZ}}}\left({M}_{500}^{{\rm{ALMA}}}\right)={Y}_{{\rm{SZ}}}\left({M}_{500}^{{\sigma }_{{\rm{v}}}}\,/{n}_{{\rm{halos}}}\right))$ is satisfied (see Extended Data Fig. 2). We note that adopting an underlying pressure model different from the A10 profile used for producing Extended Data Fig. 2 is not causing a relevant variation in the total number of subcomponents $n_{\rm halos}$, which are overall constrained to range between two to a maximum of four (except for the L15 8.5 case, resulting in $n_{\rm halos} \simeq 1$). We further note that we made several attempts in using a physically motivated subhalo mass function (for example, the theoretical model provided by Giocoli et al.[72]) instead of our simple equal-mass distribution. However, we found that removing the strong constraint of having subhalos with the same masses makes the derivation of $n_{\rm halos}$ or of the subhalo mass parameters unconstrained, resulting in unstable and heavily degenerate results. We note, in any case, that all these considerations are derived a posteriori of the SZ modelling and, thus, do not affect the significance of the reported SZ detection.

This emerging multihalo picture is also consistent with the identification reported in the literature[14,20] of double peaks in the velocity distribution. Nevertheless, we do not identify in the posterior distribution for the SZ model any notable secondary peaks that would be indicative of several pressure components in the ICM[40,41] in the direct surroundings of the Spiderweb galaxy (see however the discussion below on the results of a multicomponent analysis). This might be caused by the chance line-of-sight alignment of separate halos, but neither the spectroscopic information on the protocluster members nor the ICM constraints allow us to disentangle any distinct contributions from superimposed substructures. Similarly, any subhalos with similar masses would also be characterized by comparable

pressure distributions, in turn causing their SZ signal to be barely distinguishable. Overall, the above result suggests that the SZ effect is tracing a minor portion of the larger Spiderweb structure in which the ICM has started building up and pressurizing, whereas the rest of system, extending over scales of tens of Mpc (refs. [11,12,20,25,35,49,65]) and tracing the region encompassed by the turnaround radius of the overdensity, has yet to undergo virialization. Any SZ signals associated with further subhalos or more extended structure are not constrained by the observations, which have limited sensitivity and may suffer from large-scale interferometric filtering.

Multiple SZ components

The synchronous multiellipsoidal sampling[73] typical of the main nested sampling algorithms—and, in particular, of dynesty (refs. [45,74]), the library used for our analysis—would naturally break into separate posterior modes in the presence of several peaks in the posterior density function. This would be the case, for instance, in the presence of several halos, resulting in distinct SZ components (as reported in, for example, refs. [40,41] in the case of merging cluster systems). However, as mentioned above, we do not find such evidence in the posterior probability distribution for any of the main modelling runs presented in this work. Nevertheless, we tested for the potential presence of any extra SZ features by performing a multicomponent analysis. In particular, we consider the same model description as in the single-halo case above but introduce further SZ terms. As in the case of the radio-source modelling, to avoid label switching, we introduce an ordering condition on the centroid coordinates of the SZ components. This is applied first to the right ascension parameters and then to the declination direction, to test against any bias potentially introduced by the specific prior choice. Nevertheless, the results are found to be consistent between the two modelling sets.

Independently of the model used to describe the underlying pressure distribution, the sampler identifies a secondary SZ feature $27\,.\,{3}_{-3.8}^{+2.{4}^{{\prime\prime}}}$ southeast of the Spiderweb galaxy, corresponding to approximately 226 kpc at the protocluster redshift.

This falls right at the r_{500} boundary of the main SZ component, implying that this secondary feature, if real, would be associated with a halo distinct from the one in which the Spiderweb galaxy is embedded. The actual existence of such a structure however cannot be firmly assessed. The images produced with the high-resolution algorithm (see the ‘Sparse imaging’ section below) or after subtracting the radio-source model from the low-resolution set do not provide any clear evidence for any off-centre SZ structure. A lack of spatial correspondence is also noted with respect to the protocluster members, as the secondary SZ component cannot be clearly associated with any specific concentration of member galaxies, indicative of a distinct collapsed halo. The absence of a correspondence with protocluster galaxies further limits (if not excludes) the chances for the SZ component to be associated with the secondary velocity peak mentioned above. Above all this, the Bayesian evidence of the model comprising two SZ components improves on the one-component case by a factor of only $(\log \,{Z}_{{{\rm{n}}}_{{\rm{s}}{\rm{z}}}=2}-\log \,{Z}_{{{\rm{n}}}_{{\rm{s}}{\rm{z}}}=1})\lesssim 1.84$, corresponding to an effective significance of $\sigma_{\rm eff} \lesssim 1.90$. As such, the ALMA+ACA data available at present are not able to support the unequivocal, statistically significant identification of a secondary pressure component.

In fact, increasing the flexibility of the SZ model beyond the two-halo scenario does not provide any effective improvements in the overall reconstruction. In particular, the inclusion of a third component induces the Bayesian evidence to degrade, with a reduction with respect to the two-component case of $(\log \,{Z}_{{{\rm{n}}}_{{\rm{s}}{\rm{z}}}=3}-\log \,{Z}_{{{\rm{n}}}_{{\rm{s}}{\rm{z}}}=2})\lesssim -1.67$ and a limited improvement with respect to the reference single-halo model $(\log \,{Z}_{{{\rm{n}}}_{{\rm{s}}{\rm{z}}}=3}-\log \,{Z}_{{{\rm{n}}}_{{\rm{s}}{\rm{z}}}=1})\lesssim 0.17$ (which converts to $\sigma_{\rm eff} \lesssim 0.60$). At the same time, we observe that allowing for an ellipsoidal pressure distribution provides a marked improvement in the significance of the model ($\sigma_{\rm eff} \lesssim 4.41$). However, this concurrently makes the sampling converge to a hardly physical solution, with a plane-of-sky eccentricity $\varepsilon =0.9{6}_{-0.04}^{+0.02}$—that is, corresponding to a plane-of-sky minor axis being only 4% of the respective

major axis. This is mainly a consequence of the strong degeneracy between the mass (that is, the parameter controlling the overall amplitude and scale radius of the SZ signal) and the line-of-sight extent of the ICM distribution, governed by the eccentricity parameter ε. In fact, SZ data alone cannot provide information on the line-of-sight distribution of the optically thin ICM. We thus have to force the line-of-sight scale radius to be equal to the geometric mean of the major and minor axes of the three-dimensional ellipsoid, assumed for simplicity to lie on the plane of the sky.

Overall, the main consequence for the main SZ detection with data available at present is that a single spherically symmetric halo is sufficient to provide a statistically exhaustive description of the SZ footprint of the Spiderweb protocluster. The result of the elliptical modelling can only be interpreted as a marginal indication of an underlying morphological complexity, without however providing a conclusive and meaningful answer on the spatial properties of the SZ signal. At the same time, as demonstrated for the radio-source model, any resolvable asymmetry should be naturally traced by an ordered multicomponent model. Gaining a better understanding of the morphological properties of the forming ICM would require achieving improved quality from the observational side, in terms of both sensitivity and frequency coverage. Most importantly, though, the results reported above suggest that the robust identification of the SZ signal already with the simple spherical model provides only a lower limit to the actual significance of the detection and that this could only be enhanced when including in our models the description for any irregular and asymmetric features.

Dependence of the SZ significance on the number of point-like components

First, we note that we do not observe a substantial variation in the fluxes of the compact components between the modelling runs with and without a SZ component. This provides a straightforward test of the robustness of our model reconstruction, in particular assuring against being driven in the SZ identification by the oversubtraction of the radio source. To properly assess whether the significance of the SZ signal is, however, dependent on the specific assumption on the number of point-like components, we rerun the

SZ modelling for the entire sample of model setups considered for finding the optimal set of compact sources. A summary of such a test for our reference model (A10 UP) is provided in Extended Data Fig. 3. We note that we consider here only the case $n > 4$, as this is found to represent the minimal condition for observing a sensible improvement in the image-space residuals and for the sampler not to suffer from slow convergence.

The first outcome is that increasing the number of point-like components beyond our reference model ($n = 8$) induces a drop in the Bayesian evidence, thus not justifying any further extension of the radio source to $n \geq 9$ (second panel). This implies that, despite all the parameters remaining practically unvaried beyond the optimal set with $n = 8$, increasing the number of point-like sources to $n \geq 9$ makes the modelling incur data overfitting. On the other hand, before $n = 6$, the models exhibit a rapid increase in the overall significance in comparison with the null case, that is, the data-only run with no model components except for the cross-data calibration parameters (first panel). Concurrently, for $n \leq 5$, the right ascension and declination of the SZ centroid (fourth and fifth panels) are observed to roughly collapse on the values inferred for the secondary SZ feature identified at low significance in the multicomponent run discussed above. Corresponding to a region with low primary-beam amplitude ($\lesssim$0.50, depending on the specific array and band), this is compensated by an abrupt increase in M_{500} when moving to $n \leq 5$ (sixth panel), in turn resulting in a more extended (and thus more severely filtered) SZ signal. Despite being naively favoured by statistical reasoning on ${\sigma}_{{\rm{eff}}}^{{\rm{sz}}}$ (third panel), we note that the overall effective significance of the SZ models for $n \leq 5$ is, however, substantially lower than the stable $n \geq 6$ cases (first panel), thus limiting the validity of this reconstruction. Further, the cases $n \leq 5$ correspond to radio-source models that substantially underfit the data and fail in describing the complex morphology of the extended emission from the Spiderweb galaxy (for this, we refer to Extended Data Fig. 4). As such, the identification of such an offset SZ feature cannot be reliably associated with the actual presence of any physical component and might be induced by spurious systematic features in the visibility data.

Despite overall supporting the effectiveness of our modelling and choice for the reference model, the above results clearly highlight a marginal level of variance introduced by the different assumptions on the number of compact sources. To take this into account and to limit any bias consequent to the choice of a specific model as reference, we thus decide to use Bayesian model averaging[75,76] to generate model-marginalized profiles when comparing our reconstruction with observations (see Fig. 1). In practice, we computed cross-model posterior probability distributions by applying a weighted average to the original model posteriors, with the weight of each sampling point set equal to the respective Bayesian evidence. To fully account for any degeneracies between different model parameters, we generate radio and SZ *uv* models for each posterior sample and each pressure model, and then apply the Bayesian model averaging reduction to the resulting collection of *uv* profiles.

Systematics in the Bayesian analysis

We note that, because the dynamical mass estimates are computed assuming virial equilibrium for the entire structure, their values represent upper limits to the real mass of the Spiderweb protocluster. On the other hand, our SZ-derived mass estimates might be affected by non-trivial systematic biases associated with the assumptions used in our modelling.

First, the conversion between SZ signal and total mass is derived under hydrostatic equilibrium considerations. Given the disturbed nature of a galaxy protocluster, we can instead expect notable non-thermal contributions to the overall pressure support to the ICM—ranging from turbulent motion to dynamical effects caused by recent or continuing merger events. At the same time, hydrodynamical simulations[77,78,79] show that the growth of the mass of a system resulting from a merger event does not correspond to a concomitant increase of the thermal SZ signal, owing to the temporal offset between mass evolution and gas thermalization. If the Spiderweb protocluster is observed while experiencing a merging process, the intrinsic SZ signal would thus be lower than expected from standard mass-observable scaling relations.

Second, the reconstruction of the SZ signal relies on the assumption of self-similarity across cosmic time of the halo properties. Although this is observed[80] to hold on large scales for galaxy clusters even up to $z \simeq 1.9$, we still lack an exhaustive description of the average thermodynamic properties for the ICM within protocluster complexes, as well as any observational information for the potential self-similar appearance of protocluster halos with the ones in their $z \lesssim 2$ descendants. Also, for the sake of computational feasibility, we rely on the relatively strong assumption of spherically symmetric distribution for the electron pressure within the ICM. Nevertheless, the overall depth of the available data limits the possibility of obtaining better constraints on the SZ effect from the Spiderweb complex or achieving an improved separation of the signals from the radio source and the underlying SZ effect able to highlight any diffuse ICM halo with low surface brightness. Similarly, asymmetries in the morphology of the pressure distribution and, therefore, in the resulting SZ distribution, as well as to any secondary ICM components populating the Spiderweb complex, cannot be firmly assessed (see, however, ‘Multiple SZ components’ above for a discussion).

Third, given the limited information across the millimetre/submillimetre window with sufficient sensitivity to constrain the SZ spectrum, we are not able to disentangle any contribution to the overall SZ signal in the direction of the Spiderweb protocluster from the kinetic term. Similarly, we cannot constrain the relativistic correction to the SZ spectrum. Such effect is however generally subdominant at the virial temperature expected for such a low-mass system (for example, the relativistic correction to the thermal SZ effect at 2 keV is on the order of 1.2%; we refer to Mroczkowski et al.[3] for a review of the different contributions to the SZ effect). Therefore, we assume the measured SZ signal to be entirely caused by its non-relativistic thermal component.

Sparse imaging

The standard tools available at present for performing radio-interferometric imaging and deconvolution are not optimally suited to the joint reconstruction of the signal from an extended radio source superimposed over a diffuse SZ decrement. Cross-contamination may in fact cause both

an underestimation of the flux of the former when using CLEAN-like algorithms[81], and—at the same time—introduce a notable suppression of the underlying SZ footprint. Also, common approaches exploiting scale separation between the signature from the radio sources and the ICM in galaxy clusters[82,83] would not be able to provide a robust characterization of the two signals, as their characteristic scales exhibit a broad overlap in the case of the Spiderweb protocluster.

Building on the extensive literature on compressed sensing[84,85,86] and sparsity-based component separation[87,88] and imaging[89,90,91,92], we thus developed an algorithm for taking full advantage of both the different and, in the case of the SZ effect, well-constrained spectral behaviour of the measured signals, as well as the information on the different spatial-correlation properties. In particular, we assume the total surface brightness I_ν in a given direction on the plane of sky (x, y) and at a given frequency ν to be described as

$${I}_{\nu }(x,y)={I}_{{\rm{RS}}}(x,y)\cdot {(\nu /{\nu }_{0})}^{\alpha (x,y)}+{I}_{{\rm{SZ}}}(x,y)\cdot {g}_{{\rm{SZ}}}(\nu)$$
(1)

Here $I_{\rm RS}(x, y)$ is the surface brightness of the radio source computed at the reference frequency ν_0, whereas $\alpha(x, y)$ is the corresponding spatially varying spectral index. The terms $I_{\rm SZ}(x, y)$ and $g_{\rm SZ}(\nu)$ denote, instead, the amplitude in units of Compton y and spectral scaling for the thermal SZ effect[2,3]. Because, as already discussed above, we are not including in our SZ model any corrections owing to relativistic terms, the spectral SZ scaling is determined solely by the frequency and therefore does not contribute to the overall model through any specific free parameter. To achieve a high-fidelity reconstruction of the complex morphology of the structure of the radio jet from the Spiderweb galaxy, we model $I_{\rm RS}(x, y)$ and $\alpha(x, y)$ as a collection of pixels with angular scale matched to the longest baseline according to the Nyquist sampling theorem (given the maximum uv distance of $u_{\rm max} \simeq 1.3$ Mλ, we set the pixel scale equal to $0.5/u_{\rm max} \simeq 0.07''$). On the other hand, the signal-to-noise ratio of the SZ effect from the Spiderweb complex is not high enough to allow for adopting an analogous

approach to constrain the SZ decrement. To exploit the expected large-scale correlation, we hence describe $I_{SZ}(x, y)$ through an isothermal, circular β model[93], with free centroid coordinates, amplitude normalization and core radius, whereas we arbitrarily fix β to a value of 1 (commonly used in ICM studies and adopted, for example, by the SPT collaboration as the base for their matched-filter cluster template[56,57,94]). We tested against potential biases introduced by the specific choice of β but did not observe a dependence of the solution on this parameter. Clearly, this model is less refined than the gNFW profile broadly used in the literature and in the low-resolution analysis presented above. Nevertheless, here, we are mostly interested in capturing the average properties of the faint SZ distribution, while providing a good estimate of the base level on top of which to measure the surface brightness of the radio structure and avoid its undersubtraction when imaging the SZ effect.

Using a non-parametric approach for reconstructing the radio jet has the clear advantage of allowing for constraining the signal on the finest scales accessible in the ALMA+ACA measurements. However, this causes the model to comprise \(O(1{0}^{4})\) free parameters, making the inference problem intractable through posterior sampling methods (for example, Markov chain Monte Carlo, nested sampling). At the same time, solving simultaneously for all the free parameters in the model of equation (1) through an optimization scheme would imply facing the shortcomings of a non-convex likelihood function. We hence use an iterative, block-coordinate optimization process[95] with proximal point update[96,97,98]. In this scheme, we alternatively solve at a time only for one set of parameters—namely, the radio-source amplitude, its spatially varying spectral index and the SZ centroid coordinates, normalization and core radius—through a box-constrained limited-memory Broyden–Fletcher–Goldfarb–Shanno algorithm[99,100]. We use the implementation provided in the SCIPY[101] package and assume all the standard values for controlling the convergence of the algorithm. The selection of the specific coordinate set to solve for is performed by means of the Gauss–Southwell rule[102,103], which introduces an ordering in the optimization blocks based on the norm of the Jacobian of the corresponding likelihood function.

When modelling the radio-source amplitude, we consider a LASSO-like regression[104] step to promote sparsity,

$${L}_{{\rm{R}}{\rm{S}}}=-\,\frac{1}{2}{\chi }_{{\rm{R}}{\rm{S}}}^{2}+{\lambda }_{{\ell }_{1}}\parallel {I}_{{\rm{R}}{\rm{S}}}{\parallel }_{{\ell }_{1}}$$
(2)

in which ${\chi }_{{\rm{RS}}}^{2}$ is computed when fixing all the model parameters but I_{RS}, whereas the constant ${\lambda }_{{{\ell }}_{1}}$ controls the strength of the ℓ_1-norm regularization. To mitigate the bias and amplitude truncation[105] introduced by the ℓ_1 norm, we further iterate over the full block-coordinate loop after reweighting the regularization constant according to the previous solution for the radio source. At every main step k of the optimization algorithm, we scale the regularization hyperparameter as

$${\lambda }_{{{\ell }}_{1}}^{(k)}=\frac{{{\epsilon }}^{(k-1)}}{{{\epsilon }}^{(k-1)}+\parallel {I}_{{\rm{RS}}}^{(k-1)}{\parallel }_{{{\ell }}_{1}}}{\lambda }_{{{\ell }}_{1}}^{(k-1)}.$$
(3)

in which $\epsilon^{(k)}$ is an auxiliary coefficient aimed at ensuring numerical stability. We assume this to decrease with the number of reweighting iterations, ${{\epsilon }}^{(k)}=\max \left({{\epsilon }}^{(0)}\,1{0}^{-k},{\sigma }_{{\rm{pix}}}\right)$, with $\epsilon^{(0)}$ given by the standard deviation per pixel of the signal as measured directly from the dirty image and σ is the RMS of the image-space noise per pixel[105]. Similarly, we can express the initial value ${\lambda }_{{{\ell }}_{1}}^{(0)}$ in units of σ_{pix} as[106] ${\lambda }_{{{\ell }}_{1}}^{(0)}=n{\sigma }_{{\rm{pix}}}{\Sigma }_{i}{w}_{i}$, in which w_i are the weights of the individual visibilities and n is a user-defined quantity controlling the overall depth of the modelling process (that is, the optimization algorithm neglects any image-space features with significance lower than about $n\sigma_{pix}$). Each block-coordinate optimization loop is then run until the relative variation in the value of the likelihood function between two full sets of updates falls below

an arbitrary tolerance of 10^{-8}. The same tolerance on the relative change of the likelihood function between two consecutive solutions is further used for interrupting the iteration over the reweighting steps.

For the starting values for the model parameters, we consider ${I}_{{\rm{RS}}}^{(0)}$ to be null everywhere and set ${\alpha }_{{\rm{RS}}}^{(0)}$ to the spectral index of −0.70 typical for synchrotron emission. The coordinates of the SZ distribution are initially set on the phase centre of the Band 3 ALMA observations, roughly coinciding with the Spiderweb galaxy, whereas the initial amplitude and scale radius are initially set equal to zero and to an arbitrary scale of 30″, respectively. We tested for such choices by injecting random values for each parameter block at the beginning of the optimization run and found no substantial deviations in the final results.

In our analysis, we attempted to impose an extra smoothing to the radio-source amplitude through a total squared variation regularization[107], found to guarantee, in astronomical imaging, better results than the sole ℓ_1-norm or standard isotropic total variation. However, we observed only a minor impact on the final solution and hence decided not to include any total squared variation term in the likelihood function. Similarly, the limited significance of the available data did not allow for successfully exploiting any wavelet-based algorithms, generally demonstrated[89,106,108,109,110] to provide an improved performance compared with simple image-based techniques.

In our analysis, we adopt a conservative threshold $n = 4$. In our tests, a lower parameter showed improved subtraction capabilities but also deteriorated purity in the solution. As an internal cross-check, we further tested against the recovered flux of the radio source, finding a maximum-a-posteriori estimate from the sparse modelling of 2.606 mJy against the value $2.60{2}_{-0.007}^{+0.008}\,{\rm{mJy}}$ inferred using the nested sampling approach. Finally, the reconstructed source manifests a morphology that matches closely the signal measured by the VLA in X-band (Extended Data Fig. 5).

Insights from multiwavelength information

The combination of the ALMA+ACA data with the wealth of the multiwavelength information available on the Spiderweb galaxy allows us to propose several potential scenarios that could explain the intensity and morphology of the SZ signal and the associated offset with respect to the Spiderweb galaxy itself. Although mild spatial deviations between ICM halos and central galaxies are not a surprise even in moderately disturbed systems, the identification of such a displacement can inform our observational understanding of the dynamical state of a structure (see also ‘Comparison with simulated protoclusters’ below). Nevertheless, the available ALMA+ACA observations are not allowing to gain any insights on the physical and thermodynamic state of the forming ICM beyond the SZ detection—for example, the fraction of the total ICM reservoir that has already thermalized, the overall virialization state, the relative contribution of large-scale accretion, mergers and AGN feedback in heating the assembling ICM, the multiscale morphology of the proto-ICM and its connection (or absence thereof) with the physical properties of protocluster galaxies. Obtaining a robust characterization of the intertwined processes determining the complex physical state of the forming ICM will thus necessarily require to further improve the observational coverage of the protocluster.

Given the inherently disturbed nature of the Spiderweb protocluster, the system could be characterized by marked contributions from the kinetic SZ effect in the direction of intracluster substructures with non-negligible relative velocity with respect to the average bulk motion of the whole system. In particular, several studies[15,19] report that the protocluster galaxies in the region southwest of the Spiderweb galaxy exhibit a relative blueshift compared with that of the central galaxy. This was interpreted as evidence for the presence of an approaching background feature, potentially tracing a filament infalling onto the protocluster core. Assuming that such structure hosts a nascent intracluster gas, this would thus be associated with a reduction of the total SZ flux owing to the infilling of the thermal SZ signal by a kinetic SZ contribution. An approaching ICM component with the same mass as that corresponding to the observed SZ feature would require, to completely cancel out its own thermal SZ effect, a line-of-sight velocity of ${v}_{{\rm{LoS}}}=-\,1\,,92{8}_{-699}^{+641}\,{\rm{km}}\,{{\rm{s}}}^{-1}$ when assuming the temperature estimate \

${T}_{{\rm{e}}}=2.0{0}_{-0.40}^{+0.70}\,{\rm{keV}}$ from the Chandra X-ray data[23]. Alternatively, assuming the empirical mass–temperature scaling by Lovisari et al.[111] and the resulting temperature of ${T}_{{\rm{e}}}=0.8{5}_{-0.25}^{+0.42}\,{\rm{keV}}$, the required velocity would be ${v}_{{\rm{LoS}}}=-79{6}_{-266}^{+356}\,{\rm{km}}\,{{\rm{s}}}^{-1}$. Notably, the former velocity value roughly corresponds to the largest velocity reported by Miley et al.[19], whereas both estimates are consistent with the average velocity for both the Lyα-emitting satellite members[19] or the blueshifted components[14] measured by the Spectrograph for INtegral Field Observations in the Near Infrared[112] (SINFONI) on the Very Large Telescope (VLT). The same observations further show the presence of a small overdensity of redshifted members in the region immediately northeast of the Spiderweb galaxy. Similarly, the asymmetry in the SZ effect with respect to the central galaxy could be alternatively caused by an enhancement of the total SZ decrement owing to a negative kinetic contribution.

Although a notable contribution of the kinetic SZ effect can thus explain some of the multifrequency observations, we argue that this is probably not the only process responsible for the observed asymmetry. In fact, if the kinetic SZ effect were solely responsible for the skewed SZ decrement in the direction of the Spiderweb core, the X-ray emission should not be affected. We do instead observe a marginal asymmetry between the extended halo and the Spiderweb galaxy also in the Chandra X-ray data of the Spiderweb field, from which the bulk of the extended thermal emission is reported[23,25] to have a mild shift towards the northeast with respect to the X-ray signal from the central AGN (central panel of Extended Data Fig. 6; we refer to a forthcoming, dedicated paper (M. Lepore et al., manuscript in preparation) for an extended description of the X-ray data processing). However, a further suppression of the overall X-ray signal is observed in the region southwest of the main core—analogous to the SZ case—hence suggesting a local depletion in the forming intracluster halo. Recently, Anderson et al.[8] and Carilli et al.[9] reported direct evidence for the interaction of the extended radio jets with the Spiderweb environment, which could potentially result in the inflation of cavities in the diffuse protocluster gas. As commonly observed in the cores of local cool-core clusters, the presence of buoyant bubbles would induce a reduction of both

the X-ray and SZ signals over the limited regions covered by the cavities—provided that the bulk of their pressure is because of non-thermal components[113]—and thus contribute to the observed asymmetry in the ICM emission.

Although potentially important, none of the processes discussed above seem to effectively explain why the proto-ICM centroid is offset with respect to the Spiderweb galaxy (Fig. 3) and the centres of the Lyα halo (right panel of Extended Data Fig. 6) and of the Hα nebula[49] in which the Spiderweb galaxy is embedded. In a relatively relaxed system, the central galaxy should in fact sit deep in the centre of the gravitational well of the protocluster complex, along with most of the gas undergoing virialization, whereas the warm gas nebula is expected[8,9,47,48] to be embedded in an extended halo of hot plasma. However, the lack of strong evidence for dominant SZ or X-ray signatures from such a large-scale halo, in combination with the numerous measurements[33,63,114] showing that the central approximately 70 kpc in the Spiderweb complex hosts a massive reservoir of atomic and molecular gas, hints at a substantial, possibly predominant presence in the core region of warm and cold gas phases rather than ionized gas at the virial temperature. Emonts et al.[115] further report the detection of a tail in the CO $J = 1$–0 line emission from the surroundings of the Spiderweb galaxy, interpreted as being associated with preferential accretion of cold gas along a filament. Notably, such emission is oriented along the same direction as the asymmetry in the SZ and X-ray morphologies. At the same time, past observations provide evidence for relative motion of the Spiderweb galaxy within its own molecular halo[33,114], as well as a large-scale systematic asymmetry in the velocity distribution of the protocluster members, denoting ordered flows of galaxies towards the central galaxy[14,19]. Such motion could also naturally explain the difference observed in the radio morphology of the two opposite jets from the central radio galaxy[8,9,48] (see Fig. 3). The western lobe is in fact found to be experiencing a stronger interaction with the environment than its eastern counterpart. Although recent studies[8,9] based on VLA observations of the Spiderweb galaxy satisfactorily explained this behaviour as resulting from the inhomogeneity of the density field in the surroundings of the Spiderweb galaxy, it is not possible to exclude that the

relative motion of the radio galaxy within the diffuse intracluster atmosphere is in fact contributing to the overall ram pressure enhancement. All these hints, taken together, suggest that the shift between the bulk distribution of the proto-ICM and the Spiderweb galaxy may be the net result of the merging activity experienced by the system, with a multiphase subhalo or filament infalling into the cold, compact core from the east–northeast and causing a local enhancement of the SZ and X-ray signals through shock or adiabatic compression of the intracluster gas.

It is worth mentioning that the clear displacement between the proto-ICM halo and the Spiderweb galaxy might be indicative of the subdominant role of AGN feedback in providing the energy support required to heat the assembling intracluster gas. As reported by Tozzi et al.[23], the radio galaxy might be observed in the midst of the establishment of the radio-mode feedback regime and potentially providing a subefficient heating mechanism. We can use the estimates of the non-thermal pressure and the volume of the radio-bright regions provided by Carilli et al.[9] to roughly evaluate the non-thermal energy provided by the radio jet, $E_{\mathrm{nth}} \simeq 6 \times 10^{60}$ erg. Models of jet propagation through the ICM[116] suggest that a comparable amount should already be deposited into the shocked intracluster gas. In fact, the jet energy E_{nth} is comparable with the ICM thermal energy within the approximately 60-kpc sphere encapsulating the radio jets. This means that the mechanical energy might have a non-negligible impact on the thermodynamics of the gas (and SZ flux) from the same region. However, to effectively affect the thermal energy of the ICM in a few $10^{13}\,M_{\odot}$ cluster up to r_{500}, the jet has to operate for an order-of-magnitude longer time. All in all, despite the broad evidence[7,8,9,20,25,117] for a major cross-interaction between the central AGN and the assembling environment, the available data are still not allowing for inferring a conclusive answer on the relative contribution of AGN heating (versus, for example, gravitational process as infall and major mergers) to the overall thermal energy budget of the system. An investigation of the energetics and thermodynamics of the AGN–ICM system will thus be deferred to a future study.

Finally, it is important to note that the correspondence of the dominant features and, in particular, of the lack of thermal signal to the west of the Spiderweb galaxy in both the X-ray and SZ data (see central panel of Extended Data Fig. 6) exclude that any potential undersubtraction of the signal from the extended radio source could have a dominant effect on our results (see, for example, ‘Nested posterior sampling’ above for a collection of tests on the model reconstruction). Nevertheless, improvements to our reconstruction would require a more extensive observational coverage, including deeper X-ray and multifrequency SZ measurements, and higher angular resolution in particular for the latter. It is worth noting that the measured integrated Compton *Y* parameter—$Y(< 5{r}_{500})=(1.6{8}_{-0.32}^{+0.35})\times 1{0}^{-6}\,{{\rm{Mpc}}}^{2}$—corresponds to a SZ flux density of about 0.6 mJy at 100 GHz. Such measurements are however barely in reach of current-generation subarcminute-resolution millimetre-wave facilities.

Comparison with simulated protoclusters

The parametric modelling presented in the previous sections has the fundamental advantage of providing a straightforward and highly informative characterization of the average properties of the Spiderweb protocluster (for example, the halo mass). Nevertheless, owing to the many simplifying assumptions required to perform the model reconstruction using sensitivity-limited measurements, the analytic models used in our analysis are not able to describe the finer details of the intracluster gas. Furthermore, all the models for the ICM pressure profile available in the literature at present and explored in this work are calibrated on systems on a different mass or redshift range than the Spiderweb protocluster. In fact, we still lack an analytic description derived from observational considerations of the average thermodynamic distributions and, in particular, of the characteristic pressure profile for galaxy clusters and protocluster structures at $z \gtrsim 2$. Finally, because these models provide mean pressure profiles, they cannot capture the diversity and the possibly complex morphology of the pressure structure of the ICM in protoclusters.

The shortcomings of the parametric modelling motivates the use of hydrodynamical simulations to perform a more thorough exploration of the

properties of the SZ signal measured in the ALMA+ACA observations of the Spiderweb system. We stress again that this analysis has only the scope of presenting a qualitative comparison with numerical predictions of objects evolved in a cosmological environment, without any pretence to identifying and analysing an exact numerical equivalent of the Spiderweb complex. The simulated protoclusters analysed in this work are taken from the 'Dianoga' simulation set[18]. In particular, these comprise five regions that at $z = 0$ have sizes in the range 25–40 comoving Mpc h^{-1}, selected from a parent 1 h^{-1} Gpc^3 dark-matter-only cosmological volume around five massive objects (with M_{500} at $z = 0$ in the range $(5.636–8.811) \times 10^{14}\ M_{\odot}$ h^{-1}). Such regions are resimulated with the *N*-body hydrodynamical code OpenGadget3 at high resolution (with dark-matter and gas-particle masses of $3.4 \times 10^7\ h^{-1}\ M_{\odot}$ and $6.1 \times 10^6\ h^{-1}\ M_{\odot}$, respectively). Besides hydrodynamics, which is described by an improved version[118] of the smoothed-particle hydrodynamics, these simulations include radiative cooling, star formation, chemical enrichment and energy feedback from both supernovae and AGN (we refer to Planelles et al.[17] and Bassini et al.[18] for details on previous versions of the simulations used here and to a forthcoming paper by S. Borgani et al. (manuscript in preparation) for this specific sample).

The high-resolution regions are large enough to resolve several halos, some of which—but not all—are progenitors of the main clusters selected at $z = 0$. For the purpose of this comparison, we select all systems with a mass $M_{500} \gtrsim 1.3 \times 10^{13}\ M_{\odot}$ at $z = 2.16$, for a total sample of 27 distinct objects. SZ maps are generated at the corresponding snapshot using the map-making software SMAC (ref. [119]) and projected onto the visibility plane. We note that, to simplify the comparison, the projected SZ maps are generated considering only the thermal component and they do not encompass any kinetic SZ term. Despite the obvious discrepancies associated with the specific differences in the morphological properties and dynamical states of the simulated protoclusters with respect to the Spiderweb complex, Fig. 2 shows that the mock *uv* profiles for simulated halos with mass in the same range of the parametric estimates (approximately $3 \times 10^{13}\ M_{\odot}$) are broadly consistent with the ALMA+ACA measurements.

To further gain visual insight into the simulated protoclusters, we produce synthetic maps by projecting the simulated ALMA+ACA visibilities back into the image space, consistent with the imaging-reconstruction process of Extended Data Fig. 6. Three out of the full sample of simulated clusters are shown in Extended Data Fig. 7, with one specifically selected as exhibiting a SZ footprint resembling that observed for the Spiderweb protocluster, one as the most massive halo in the simulated set and another as exhibiting a displacement between the SZ peak and the halo centre analogous to what is observed between the SZ signal and the central galaxy in the Spiderweb complex. We checked that, for the entire sample, the physical displacement between the dark-matter halos and the corresponding dominant galaxies were never larger than about 6 kpc, more than an order of magnitude smaller than the observed SZ-galaxy displacement in the Spiderweb protocluster. In the context of the discussion on such an offset, the halo centroid and galaxy for each of the simulated halos can thus be considered to be consistent. To introduce realistic noise and optimally reflect the noise properties of the available observations, we jackknife the original measurements by changing the sign of a randomly selected half of the visibilities. This noise-like realization is finally added to the idealized SZ data. Consistent with the *uv* result above, we can observe an overall agreement in terms of the average amplitude of the SZ surface brightness in the reconstructed images between the SZ signal measured in the ALMA+ACA data (Extended Data Fig. 6) and the estimate for the simulated low-mass system. We note that the selection of a specific projection axis has only a minor effect on the measured SZ amplitude. Over the whole sample, the change in both the peak and the average SZ signal associated with a different projection exhibits a RMS variation of about 10%, reaching a maximum of around 30% in only three cases, corresponding to objects that manifest morphological disturbances.

In any case, what is most important in such a comparison is the possibility of gaining a sensible understanding of the effect of radio-interferometric filtering on the observed distribution of the SZ signal. In the specific case shown in the bottom row of Extended Data Fig. 7, we can observe that the displacement between the SZ distribution and the centre of the corresponding dark-matter halo (in turn roughly coincident with the position of the most massive galaxy in the structure; see above) is

exaggerated by the radio-interferometric suppression of the bulk, relatively symmetric SZ signal on large scales. The net result is an enhancement of local SZ asymmetries that finally drive the definition of the observed ICM morphology. In fact, the observed offset is not indicative of an absolute displacement between the ICM and the Spiderweb galaxy but rather of a positional mismatch between the peak of the pressure distribution and the radio galaxy, typical of disturbed systems.

It is clear that, to move this comparison beyond the sole qualitative analysis and to perform a robust cross-characterization of the observed and simulated SZ signals, a broader exploration of the model parameter space in the available hydrodynamical simulations would be required. Such a detailed investigation will however be deferred to a future work.

Data availability

This work is based on ALMA and ACA measurements (project codes: 2015.1.00851.S, 2016.2.00048.S, 2018.1.01526.S) publicly available at http://almascience.org/aq/. The Chandra observations can be obtained from the CDA (http://cxc.cfa.harvard.edu/cda/). Access to the VLT/FORS1 images is possible through http://archive.eso.org/eso/eso_archive_main.html. The Hubble Space Telescope image shown in this work is publicly available at https://archive.stsci.edu/. At the moment, technical limitations do not allow us to guarantee safe, stable and efficient public access to the full simulation set. As such, on request, the corresponding author will provide the data products from the Dianoga cosmological hydrodynamical simulations used for the comparisons presented in this manuscript. Source data for all figures are instead provided with this paper.

Code availability

All the analyses in this work rely on the following Python packages: numpy[120], scipy[101], matplotlib[121], aplpy[122,123]. The nested sampling is performed using dynesty[45,74]. We use astropy[124,125] to manage all the common astronomical tasks and the just-in-time compiler and automatic

differentiation in jax[126] to improve the performance of the optimizations in the sparse-image reconstruction. The sparse-imaging problem is solved by means of the L-BFGS-B implementation provided in the minimize routine that is part of the scipy.optimize subpackage. To compute non-uniform fast Fourier transforms of the ALMA+ACA data, we use the Python interface to finufft[42]. Owing to limited readability, documentation and user-friendliness of their interfaces, the full analysis scripts are not yet ready for public release. We are, however, keen on collaborating with anyone willing to use the modelling tools, available from the corresponding author on request.

References

1. Kravtsov, A. V. & Borgani, S. Formation of galaxy clusters. *Annu. Rev. Astron. Astrophys.* **50**, 353–409 (2012).

2. Sunyaev, R. A. & Zeldovich, Y. B. The observations of relic radiation as a test of the nature of X-ray radiation from the clusters of galaxies. *Comments Astrophys. Space Phys.* **4**, 173 (1972).

3. Mroczkowski, T. et al. Astrophysics with the spatially and spectrally resolved Sunyaev-Zeldovich effects: a millimetre/submillimetre probe of the warm and hot universe. *Space Sci. Rev.* **215**, 17 (2019).

4. Overzier, R. A. The realm of the galaxy protoclusters: a review. *Astron. Astrophys. Rev.* **24**, 14 (2016).

5. Zhang, C., Yu, Q. & Lu, Y. A baryonic effect on the merger timescale of galaxy clusters. *Astrophys. J.* **820**, 85 (2016).

6. Carilli, C. L., Röttgering, H. J. A., van Ojik, R., Miley, G. K. & van Breugel, W. J. M. Radio continuum imaging of high-redshift radio galaxies. *Astrophys. J. Suppl. Ser.* **109**, 1–44 (1997).

7. Pentericci, L., Roettgering, H. J. A., Miley, G. K., Carilli, C. L. & McCarthy, P. The radio galaxy 1138–262 at z=2.2: a giant elliptical galaxy at the center of a proto-cluster? *Astron. Astrophys.* **326**, 580–596 (1997).

8. Anderson, C. S. et al. The Spiderweb protocluster is being magnetized by its central radio jet. *Astrophys. J.* **937**, 45 (2022).

9. Carilli, C. L. et al. X-ray emission from the jets and lobes of the Spiderweb. *Astrophys. J.* **928**, 59 (2022).

10. Pratt, G. W. et al. The galaxy cluster mass scale and its impact on cosmological constraints from the cluster population. *Space Sci. Rev.* **215**, 25 (2019).

11. Kurk, J. D., Pentericci, L., Röttgering, H. J. A. & Miley, G. K. A search for clusters at high redshift. III. Candidate Hα emitters and EROs in the PKS 1138–262 proto-cluster at $z = 2.16$. *Astron. Astrophys.* **428**, 793–815 (2004).

12. Kurk, J. D., Pentericci, L., Overzier, R. A., Röttgering, H. J. A. & Miley, G. K. A search for clusters at high redshift. IV. Spectroscopy of Hα emitters in a proto-cluster at $z = 2.16$. *Astron. Astrophys.* **428**, 817–821 (2004).

13. Saro, A. et al. Simulating the formation of a protocluster at $z \sim 2$. *Mon. Not. R. Astron. Soc.* **392**, 795–800 (2009).

14. Kuiper, E. et al. A SINFONI view of flies in the Spiderweb: a galaxy cluster in the making. *Mon. Not. R. Astron. Soc.* **415**, 2245–2256 (2011).

15. Shimakawa, R. et al. Identification of the progenitors of rich clusters and member galaxies in rapid formation at $z > 2$. *Mon. Not. R. Astron. Soc.* **441**, L1–L5 (2014).

16. Di Mascolo, L., Churazov, E. & Mroczkowski, T. A joint ALMA–Bolocam–*Planck* SZ study of the pressure distribution in RX J1347.5–1145. *Mon. Not. R. Astron. Soc.* **487**, 4037–4056 (2019).

17. Planelles, S. et al. Pressure of the hot gas in simulations of galaxy clusters. *Mon. Not. R. Astron. Soc.* **467**, 3827–3847 (2017).

18. Bassini, L. et al. The DIANOGA simulations of galaxy clusters: characterising star formation in protoclusters. *Astron. Astrophys.* **642**, A37 (2020).

19. Miley, G. K. et al. The Spiderweb galaxy: a forming massive cluster galaxy at $z \sim 2$. *Astrophys. J.* **650**, L29–L32 (2006).

20. Pentericci, L. et al. A search for clusters at high redshift. II. A proto cluster around a radio galaxy at $z = 2.16$. *Astron. Astrophys.* **361**, L25–L28 (2000).

21. Hatch, N. A., Overzier, R. A., Röttgering, H. J. A., Kurk, J. D. & Miley, G. K. Diffuse UV light associated with the Spiderweb Galaxy: evidence for *in situ* star formation outside galaxies. *Mon. Not. R. Astron. Soc.* **383**, 931–942 (2008).

22. Carilli, C. L. et al. An X-ray cluster at redshift 2.156? *Astrophys. J.* **494**, L143–L146 (1998).

23. Tozzi, P. et al. The 700 ks *Chandra* Spiderweb Field. II. Evidence for inverse-Compton and thermal diffuse emission in the Spiderweb galaxy. *Astron. Astrophys.* **667**, A134 (2022).

24. Remus, R.-S., Dolag, K. & Dannerbauer, H. The young and the wild: what happens to protoclusters forming at $z = 4$? Preprint at https://arxiv.org/abs/2208.01053 (2022)

25. Tozzi, P. et al. The 700 ks *Chandra* Spiderweb Field. I. Evidence for widespread nuclear activity in the protocluster. *Astron. Astrophys.* **662**, A54 (2022).

26. Overzier, R. A. et al. On the X-ray emission of $z \sim 2$ radio galaxies: IC scattering of the CMB and no evidence for fully-formed potential wells. *Astron. Astrophys.* **433**, 87–100 (2005).

27. Champagne, J. B. et al. Comprehensive gas characterization of a $z = 2.5$ protocluster: a cluster core caught in the beginning of virialization? *Astrophys. J.* **913**, 110 (2021).

28. Arnaud, M. et al. The universal galaxy cluster pressure profile from a representative sample of nearby systems (REXCESS) and the Y_{SZ} – M_{500} relation. *Astron. Astrophys.* **517**, A92 (2010).

29. Le Brun, A. M. C., McCarthy, I. G. & Melin, J.-B. Testing Sunyaev–Zel'dovich measurements of the hot gas content of dark matter haloes using synthetic skies. *Mon. Not. R. Astron. Soc.* **451**, 3868–3881 (2015).

30. Kurk, J. D. et al. A search for clusters at high redshift. I. Candidate Lyα emitters near 1138–262 at $z = 2.2$. *Astron. Astrophys.* **358**, L1–L4 (2000).

31. Wootten, A. & Thompson, A. R. The Atacama Large Millimeter/Submillimeter Array. *Proc. IEEE* **97**, 1463–1471 (2009).

32. Iguchi, S. et al. The Atacama Compact Array (ACA). *Publ. Astron. Soc. Jpn.* **61**, 1–12 (2009).

33. Emonts, B. H. C. et al. Giant galaxy growing from recycled gas: ALMA maps the circumgalactic molecular medium of the Spiderweb in [C I]. *Mon. Not. R. Astron. Soc. Lett.* **477**, L60–L65 (2018).

34. Tadaki, K.-i. et al. Environmental impacts on molecular gas in protocluster galaxies at $z \sim 2$. *Publ. Astron. Soc. Jpn.* **71**, 40 (2019).

35. Jin, S. et al. COALAS. I. ATCA CO(1–0) survey and luminosity function in the Spiderweb protocluster at $z = 2.16$. *Astron. Astrophys.* **652**, A11 (2021).

36. McMullin, J. P., Waters, B., Schiebel, D., Youn, W. & Golap, K. CASA architecture and applications. *ASP Conf. Ser.* **376**, 127 (2006).

37. Hatziminaoglou, E. et al. The European ALMA Regional Centre Network: a geographically distributed user support model. *Messenger* **162**, 24–29 (2015).

38. Petry, D. et al. ALMA data quality assurance and the products it delivers – the contribution of the European ARC. *Messenger* **181**, 16–20 (2020).

39. Di Mascolo, L. et al. An ALMA+ACA measurement of the shock in the Bullet Cluster. *Astron. Astrophys.* **628**, A100 (2019).

40. Di Mascolo, L. et al. The massive and distant clusters of WISE Survey. SZ effect of verification with the Atacama Compact Array – Localization and Cluster Analysis. *Astron. Astrophys.* **638**, A70 (2020).

41. Di Mascolo, L. et al. Multiwavelength view of SPT-CL J2106-5844. The radio galaxies and the thermal and relativistic plasmas in a massive galaxy cluster merger at $z \simeq 1.13$. *Astron. Astrophys.* **650**, A153 (2021).

42. Barnett, A. H., Magland, J. & af Klinteberg, L. A parallel nonuniform fast Fourier transform library based on an "exponential of semicircle" kernel. *SIAM J. Sci. Comput.* **41**, C479–C504 (2019).

43. Skilling, J. in *Bayesian Inference and Maximum Entropy Methods in Science and Engineering* (eds Fischer, R., Dose, V., Preuss, R. & von Toussaint, U.) 395–405 (AIP, 2004).

44. Ashton, G. et al. Nested sampling for physical scientists. *Nat. Rev. Methods Primers* **2**, 39 (2022).

45. Speagle, J. S. DYNESTY: a dynamic nested sampling package for estimating Bayesian posteriors and evidences. *Mon. Not. R. Astron. Soc.* **493**, 3132–3158 (2020).

46. Bellman, R. & Kalaba, R. A mathematical theory of adaptive control processes. *Proc. Natl Acad. Sci.* **45**, 1288–1290 (1959).

47. Pentericci, L. et al. *HST* images of the extremely clumpy radio galaxy 1138–262 at $z = 2.2$. *Astrophys. J.* **504**, 139–146 (1998).

48. Carilli, C. L. et al. The X-ray-radio alignment in the z = 2.2 radio galaxy PKS 1138–262. *Astrophys. J.* **567**, 781–789 (2002).

49. Shimakawa, R. et al. MAHALO Deep Cluster Survey II. Characterizing massive forming galaxies in the Spiderweb protocluster at $z = 2.2$. *Mon. Not. R. Astron. Soc.* **481**, 5630–5650 (2018).

50. Higson, E., Handley, W., Hobson, M. & Lasenby, A. Bayesian sparse reconstruction: a brute-force approach to astronomical imaging and machine learning. *Mon. Not. R. Astron. Soc.* **483**, 4828–4846 (2019).

51. Handley, W. J., Hobson, M. P. & Lasenby, A. N. POLYCHORD: next-generation nested sampling. *Mon. Not. R. Astron. Soc.* **453**, 4384–4398 (2015).

52. Nagai, D., Kravtsov, A. V. & Vikhlinin, A. Effects of galaxy formation on thermodynamics of the intracluster medium. *Astrophys. J.* **668**, 1–14 (2007).

53. Böhringer, H. et al. The representative XMM-Newton cluster structure survey (REXCESS) of an X-ray luminosity selected galaxy cluster sample. *Astron. Astrophys.* **469**, 363–377 (2007).

54. McDonald, M. et al. The redshift evolution of the mean temperature, pressure, and entropy profiles in 80 SPT-selected galaxy clusters. *Astrophys. J.* **794**, 67 (2014).

55. Carlstrom, J. E. et al. The 10 meter South Pole Telescope. *Publ. Astron. Soc. Pac.* **123**, 568 (2011).

56. Bleem, L. E. et al. Galaxy clusters discovered via the Sunyaev–Zel'dovich effect in the 2500-square-degree SPT-SZ survey. *Astrophys. J. Suppl. Ser.* **216**, 27 (2015).

57. Bleem, L. E. et al. The SPTPoL extended cluster survey. *Astrophys. J. Suppl. Ser.* **247**, 25 (2020).

58. Le Brun, A. M. C., McCarthy, I. G., Schaye, J. & Ponman, T. J. Towards a realistic population of simulated galaxy groups and clusters. *Mon. Not. R. Astron. Soc.* **441**, 1270–1290 (2014).

59. Fabian, A. C. Observational evidence of active galactic nuclei feedback. *Annu. Rev. Astron. Astrophys.* **50**, 455–489 (2012).

60. Harrison, C. M. Impact of supermassive black hole growth on star formation. *Nat. Astron.* **1**, 0165 (2017).

61. Gupta, N., Saro, A., Mohr, J. J., Dolag, K. & Liu, J. SZE observables, pressure profiles and centre offsets in Magneticum simulation galaxy clusters. *Mon. Not. R. Astron. Soc.* **469**, 3069–3087 (2017).

62. Doherty, M. et al. Optical and near-IR spectroscopy of candidate red galaxies in two $z \sim 2.5$ proto-clusters. *Astron. Astrophys.* **509**, A83 (2010).

63. Emonts, B. H. C. et al. Molecular gas in the halo fuels the growth of a massive cluster galaxy at high redshift. *Science* **354**, 1128–1130 (2016).

64. Pérez-Martínez, J. M. et al. Signs of environmental effects on star-forming galaxies in the Spiderweb protocluster at $z = 2.16$. *Mon. Not. R. Astron. Soc.* **518**, 1707–1734 (2023).

65. Dannerbauer, H. et al. An excess of dusty starbursts related to the Spiderweb galaxy. *Astron. Astrophys.* **570**, A55 (2014).

66. Dannerbauer, H. et al. The implications of the surprising existence of a large, massive CO disk in a distant protocluster. *Astron. Astrophys.* **608**, A48 (2017).

67. Mantz, A. B. et al. The XXL Survey. XVII. X-ray and Sunyaev-Zel'dovich properties of the redshift 2.0 galaxy cluster XLSSC 122. *Astron. Astrophys.* **620**, A2 (2018).

68. Gobat, R. et al. Sunyaev-Zel'dovich detection of the galaxy cluster Cl J1449+0856 at $z = 1.99$: the pressure profile in *uv* space. *Astron. Astrophys.* **629**, A104 (2019).

69. Saro, A., Mohr, J. J., Bazin, G. & Dolag, K. Toward unbiased galaxy cluster masses from line-of-sight velocity dispersions. *Astrophys. J.* **772**, 47 (2013).

70. Springel, V. et al. Simulations of the formation, evolution and clustering of galaxies and quasars. *Nature* **435**, 629–636 (2005).

71. Ragagnin, A., Saro, A., Singh, P. & Dolag, K. Cosmology dependence of halo masses and concentrations in hydrodynamic simulations. *Mon. Not. R. Astron. Soc.* **500**, 5056–5071 (2021).

72. Giocoli, C., Tormen, G. & van den Bosch, F. C. The population of dark matter subhaloes: mass functions and average mass-loss rates. *Mon. Not. R. Astron. Soc.* **386**, 2135–2144 (2008).

73. Feroz, F., Hobson, M. P. & Bridges, M. MULTINEST: an efficient and robust Bayesian inference tool for cosmology and particle physics. *Mon. Not. R. Astron. Soc.* **398**, 1601–1614 (2009).

74. Koposov, S. et al. joshspeagle/dynesty: v2.0.1. https://doi.org/10.5281/zenodo.7215695 (2022).

75. Hoeting, J. A., Madigan, D., Raftery, A. E. & Volinsky, C. T. Bayesian model averaging: a tutorial. *Stat. Sci.* **14**, 382–401 (1999).

76. Fragoso, T. M., Bertoli, W. & Louzada, F. Bayesian model averaging: a systematic review and conceptual classification. *Int. Stat. Rev.* **86**, 1–28 (2018).

77. Poole, G. B. et al. The impact of mergers on relaxed X-ray clusters – II. Effects on global X-ray and Sunyaev-Zel'dovich properties and their scaling relations. *Mon. Not. R. Astron. Soc.* **380**, 437–454 (2007).

78. Rasia, E. et al. Scaling relation in two situations of extreme mergers. *Astrophys. J.* **729**, 45 (2011).

79. Krause, E., Pierpaoli, E., Dolag, K. & Borgani, S. Merger-induced scatter and bias in the cluster mass–Sunyaev–Zel'dovich effect scaling relation. *Mon. Not. R. Astron. Soc.* **419**, 1766–1779 (2012).

80. McDonald, M. et al. The remarkable similarity of massive galaxy clusters from $z \sim 0$ to $z \sim 1.9$. *Astrophys. J.* **843**, 28 (2017).

81. Thompson, A. R., Moran, J. M. & Swenson, G. W. J. *Interferometry and Synthesis in Radio Astronomy* 3rd edn (Springer, 2017).

82. Kitayama, T. et al. The Sunyaev–Zel'dovich effect at 5″: RX J1347.5–1145 imaged by ALMA. *Publ. Astron. Soc. Jpn.* **68**, 88 (2016).

83. Kitayama, T. et al. Deeply cooled core of the Phoenix galaxy cluster imaged by ALMA with the Sunyaev–Zel'dovich effect. *Publ. Astron. Soc. Jpn.* **72**, 33 (2020).

84. Donoho, D. Compressed sensing. *IEEE Trans. Inf. Theory* **52**, 1289–1306 (2006).

85. Candès, E. J., Romberg, J. K. & Tao, T. Stable signal recovery from incomplete and inaccurate measurements. *Commun. Pure Appl. Math.* **59**, 1207–1223 (2006).

86. Wiaux, Y., Jacques, L., Puy, G., Scaife, A. M. M. & Vandergheynst, P. Compressed sensing imaging techniques for radio interferometry. *Mon. Not. R. Astron. Soc.* **395**, 1733–1742 (2009).

87. Bobin, J., Rapin, J., Larue, A. & Starck, J.-L. Sparsity and adaptivity for the blind separation of partially correlated sources. *IEEE Trans. Signal Process.* **63**, 1199–1213 (2015).

88. Jiang, M., Bobin, J. & Starck, J.-L. Joint multichannel deconvolution and blind source separation. *SIAM J. Imaging Sci.* **10**, 1997–2021 (2017).

89. Carrillo, R. E., McEwen, J. D. & Wiaux, Y. Sparsity Averaging Reweighted Analysis (SARA): a novel algorithm for radio-interferometric imaging. *Mon. Not. R. Astron. Soc.* **426**, 1223–1234 (2012).

90. Honma, M., Akiyama, K., Uemura, M. & Ikeda, S. Super-resolution imaging with radio interferometry using sparse modeling. *Publ. Astron. Soc. Jpn.* **66**, 95 (2014).

91. Akiyama, K. et al. Superresolution full-polarimetric imaging for radio interferometry with sparse modeling. *Astron. J.* **153**, 159 (2017).

92. Chael, A. A. et al. Interferometric imaging directly with closure phases and closure amplitudes. *Astrophys. J.* **857**, 23 (2018).

93. Cavaliere, A. & Fusco-Femiano, R. X-rays from hot plasma in clusters of galaxies. *Astron. Astrophys.* **49**, 137–144 (1976).

94. Huang, N. et al. Galaxy clusters selected via the Sunyaev–Zel'dovich effect in the SPTpol 100-square-degree survey. *Astron. J.* **159**, 110 (2020).

95. Nesterov, Y. Efficiency of coordinate descent methods on huge-scale optimization problems. *SIAM J. Optim.* **22**, 341–362 (2012).

96. Auslender, A. Asymptotic properties of the Fenchel dual functional and applications to decomposition problems. *J. Optim. Theory Appl.* **73**, 427–449 (1992).

97. Grippo, L. & Sciandrone, M. On the convergence of the block nonlinear Gauss–Seidel method under convex constraints. *Oper. Res. Lett.* **26**, 127–136 (2000).

98. Razaviyayn, M., Hong, M. & Luo, Z.-Q. A unified convergence analysis of block successive minimization methods for nonsmooth optimization. *SIAM J. Optim.* **23**, 1126–1153 (2013).

99. Byrd, R. H., Lu, P., Nocedal, J. & Zhu, C. A limited memory algorithm for bound constrained optimization. *SIAM J. Sci. Comput.* **16**, 1190–1208 (1995).

100. Zhu, C., Byrd, R. H., Lu, P. & Nocedal, J. Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization. *ACM Trans. Math. Softw.* **23**, 550–560 (1997).

101. Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).

102. Luo, Z. Q. & Tseng, P. On the convergence of the coordinate descent method for convex differentiable minimization. *J. Optim. Theory Appl.* **72**, 7–35 (1992).

103. Nutini, J., Schmidt, M., Laradji, I. H., Friedlander, M. & Koepke, H. in *Proc. 32nd International Conference on Machine Learning* Vol. 37 (eds Bach, F. & Blei, D.) 1632–1641 (PMLR, 2015).

104. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. B Stat. Methodol.* **58**, 267–288 (1996).

105. Candès, E. J., Wakin, M. B. & Boyd, S. P. Enhancing sparsity by reweighted ℓ_1 minimization. *J. Fourier Anal. Appl.* **14**, 877–905 (2008).

106. Garsden, H. et al. LOFAR sparse image reconstruction. *Astron. Astrophys.* **575**, A90 (2015).

107. Kuramochi, K. et al. Superresolution interferometric imaging with sparse modeling using total squared variation: application to imaging the black hole shadow. *Astrophys. J.* **858**, 56 (2018).

108. Starck, J.-L., Fadili, J. & Murtagh, F. The undecimated wavelet decomposition and its reconstruction. *IEEE Trans. Image Process.* **16**, 297–309 (2007).

109. Li, F., Cornwell, T. J. & de Hoog, F. The application of compressive sampling to radio astronomy. I. Deconvolution. *Astron. Astrophys.* **528**, A31 (2011).

110. Dabbech, A. et al. MORESANE: MOdel REconstruction by Synthesis-ANalysis Estimators. A sparse deconvolution algorithm for radio interferometric imaging. *Astron. Astrophys.* **576**, A7 (2015).

111. Lovisari, L., Ettori, S., Gaspari, M. & Giles, P. A. Scaling properties of galaxy groups. *Universe* **7**, 139 (2021).

112. Eisenhauer, F. et al. SINFONI - integral field spectroscopy at 50 milli-arcsecond resolution with the ESO VLT. *Proc. SPIE* **4842**, 1548–1561 (2003).

113. McNamara, B. R. & Nulsen, P. E. J. Mechanical feedback from active galactic nuclei in galaxies, groups and clusters. *New J. Phys.* **14**, 055023 (2012).

114. De Breuck, C. et al. Feeding the spider with carbon. [CII] emission from the circumgalactic medium and active galactic nucleus. *Astron. Astrophys.* **658**, L2 (2022).

115. Emonts, B. H. C. et al. CO(1–0) detection of molecular gas in the massive Spiderweb Galaxy ($z = 2$). *Mon. Not. R. Astron. Soc.* **430**, 3465–3471 (2013).

116. Gaibler, V., Krause, M. & Camenzind, M. Very light magnetized jets on large scales – I. Evolution and magnetic fields. *Mon. Not. R. Astron. Soc.* **400**, 1785–1802 (2009).

117. Gullberg, B. et al. ALMA finds dew drops in the dusty spider's web. *Astron. Astrophys.* **591**, A73 (2016).

118. Beck, A. M. et al. An improved SPH scheme for cosmological simulations. *Mon. Not. R. Astron. Soc.* **455**, 2110–2130 (2016).

119. Dolag, K., Hansen, F. K., Roncarelli, M. & Moscardini, L. The imprints of local superclusters on the Sunyaev-Zel'dovich signals and their detectability with *Planck*. *Mon. Not. R. Astron. Soc.* **363**, 29–39 (2005).

120. Harris, C. R. et al. Array programming with NumPy. *Nature* **585**, 357–362 (2020).

121. Hunter, J. D. Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).

122. Robitaille, T. & Bressert, E. APLpy: Astronomical Plotting Library in Python. Astrophysics Source Code Library, record ascl:1208.017 (2012).

123. Robitaille, T. APLpy v2.0: The Astronomical Plotting Library in Python. https://doi.org/10.5281/zenodo.2567476 (2019).

124. The Astropy Collaboration. Astropy: a community Python package for astronomy. *Astron. Astrophys.* **558**, A33 (2013).

125. The Astropy Collaboration. The Astropy Project: building an open-science project and status of the v2.0 core package. *Astron. J.* **156**, 123 (2018).

126. Bradbury, J. et al. JAX: composable transformations of Python+NumPy programs version 0.2.5. http://github.com/google/jax (2018).

127. Appenzeller, I. et al. Successful commissioning of FORS1 – the first optical instrument on the VLT. *Messenger* **94**, 1–6 (1998).

Acknowledgements

This paper makes use of the following ALMA data: ADS/JAO.ALMA#2015.1.00851.S, ADS/JAO.ALMA#2016.2.00048.S, ADS/JAO.ALMA#2018.1.01526.S. ALMA is a partnership of ESO (representing its member states), NSF (USA) and NINS (Japan), together

with NRC (Canada), MOST and ASIAA (Taiwan) and KASI (Republic of Korea), in cooperation with the Republic of Chile. The Joint ALMA Observatory is operated by ESO, AUI/NRAO and NAOJ. This work is further based in part on observations collected at the European Southern Observatory under ESO programme 63.O-0477(A). L.D.M., A.S. and M.P. are supported by the ERC-StG 'ClustersXCosmo' grant agreement 716762. A.S. is further supported by FARE-MIUR grant 'ClustersXEuclid' R165SBKTMA. A.S. and S.B. acknowledge partial financial support from the 'InDark' INFN grant. L.D.M. and P.T. acknowledge financial contribution from the agreement ASI-INAF n.2017-14-H.0. C.L.C. acknowledges support through CXC grant G09-2-1103X. H.D. acknowledges financial support from the Agencia Estatal de Investigación del Ministerio de Ciencia e Innovación (AEI-MCINN) under grant 'La evolución de los cíumulos de galaxias desde el amanecer hasta el mediodía cósmico' with reference PID2019-105776GB-I00/DOI:10.13039/501100011033 and acknowledges support from the ACIISI, Consejería de Economía, Conocimiento y Empleo del Gobierno de Canarias and the European Regional Development Fund (ERDF) under grant with reference PROID2020010107. M.N. acknowledges INAF-1.05.01.86.20. F.R. acknowledges support from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 847523 'INTERACTIONS' and the Cosmic Dawn Center, which is financed by the Danish National Research Foundation under grant no. 140.

Author information

Authors and Affiliations

1. Astronomy Unit, Department of Physics, University of Trieste, Trieste, Italy

   Luca Di Mascolo, Alexandro Saro, Stefano Borgani & Maurilio Pannella

2. INAF - Osservatorio Astronomico di Trieste, Trieste, Italy

   Luca Di Mascolo, Alexandro Saro, Stefano Borgani, Elena Rasia, Mario Nonino & Maurilio Pannella

3. IFPU - Institute for Fundamental Physics of the Universe, Trieste, Italy

   Luca Di Mascolo, Alexandro Saro, Stefano Borgani, Elena Rasia & Maurilio Pannella

4. INFN - Sezione di Trieste, Trieste, Italy

   Alexandro Saro & Stefano Borgani

5. European Southern Observatory (ESO), Garching, Germany

   Tony Mroczkowski & Michele Ginolfi

6. Max-Planck-Institut für Astrophysik (MPA), Garching, Germany

   Eugene Churazov

7. Space Research Institute, Moscow, Russia

   Eugene Churazov

8. INAF - Osservatorio Astrofisico di Arcetri, Firenze, Italy

Paolo Tozzi

9. Instituto de Astrofísica de Canarias (IAC), La Laguna, Tenerife, Spain

Helmut Dannerbauer

10. Departamento de Astrofísica, Universidad de La Laguna, La Laguna, Tenerife, Spain

Helmut Dannerbauer

11. Argelander Institute for Astronomy, University of Bonn, Bonn, Germany

Kaustuv Basu

12. National Radio Astronomy Observatory, Socorro, NM, USA

Christopher L. Carilli

13. Dipartimento di Fisica e Astronomia, Università di Firenze, Sesto Fiorentino, Italy

Michele Ginolfi

14. Leiden Observatory, Leiden University, Leiden, The Netherlands

George Miley

15. INAF - Osservatorio Astronomico di Roma, Monte Porzio Catone, Italy

Laura Pentericci

16. Cosmic Dawn Center (DAWN), Copenhagen, Denmark

Francesca Rizzo

17. Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark

Francesca Rizzo

Contributions

L.D.M. coordinated the research project, developed the radio-interferometric imaging and modelling algorithms, performed the data analyses, planned data visualization and produced the corresponding figures, and drafted and edited the manuscript. A.S. conceived and led the ALMA+ACA proposal 2018.1.01526.S on whose data this work is based. T.M. provided substantial contributions to the interpretation of the ALMA+ACA observations and to editing the manuscript. S.B. and E.R. produced and postprocessed the ‘Dianoga’ suite of cosmological hydrodynamical simulations and contributed to writing the corresponding Methods section. E.C. provided guidance on the multihalo analysis and on the interpretation of the SZ results in the multiwavelength context. P.T. led the Chandra proposal and analysed and processed the corresponding observations. H.D. and K.B. contributed to shaping the original idea at the base of the ALMA+ACA observing programme on which this work is focused. H.D. further contributed to the interpretation of the radio-source modelling in relation to the observational properties of the protocluster galaxies. C.L.C. calibrated, processed and imaged the VLA data. M.N. performed the reprocessing of the optical VLT/FORS1 observations. M.G., G.M., M.P., L.P. and F.R. contributed to the interpretation of the multiphase gas properties presented in the ‘Insights from multiwavelength information’ section. All authors contributed substantially to discussing the results and preparing the manuscript.

Corresponding author

Correspondence to Luca Di Mascolo.

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Images from the nested sampling reconstruction.

Raw image (top) from the Band 3 ALMA+ACA data and residuals (bottom) after subtraction of the marginalized model for the central radio source obtained from the nested sampling. We stress that we are not subtracting the model component for the SZ effect when generating the residual image but the limited sensitivity is making any SZ features non-obvious. Because our analyses are performed in the native visibility space, we do not perform any cleaning step. The solid and dashed white contours in the top figure trace the data image and the radio-source model, respectively. As evident from the comparison of the two sets of contours, the model map is hardly distinguishable from the data. The circles in the top panel denote the position of the point-like model components used to described the extended signal from the Spiderweb galaxy and the continuum emission from the western protocluster members. We label the point-like sources that are observed to correspond to distinct features or known astrophysical sources

(see discussion in ‘Results’ or Extended Data Table 1). The black contours overlaid on the residual map denote the reference A10 UP model for the SZ signal. The SZ component manifests a clear shift with respect to the Spiderweb galaxy, consistent with the result from the independent sparse-imaging analysis (Extended Data Fig. 6). Consistent with the nested sampling analysis, the images are produced after applying an upper *uv* cut of 65 kλ to the input data. As for Extended Data Fig. 6, the triangle marks the position of the Spiderweb galaxy. All the contour levels are arbitrary and chosen with the sole intent of optimally marking the position of the best-fit SZ model.

Extended Data Fig. 2 Integrated SZ flux as a function of the number of subhalos.

The total dynamical mass ${M}_{500}^{{\sigma }_{{\rm{v}}}}$ is here computed assuming a velocity dispersion of $\sigma_{\rm v} = 683\ {\rm km\ s^{-1}}$ as reported by Shimakawa et al.[15] for a 0.53-Mpc region around the Spiderweb galaxy. The green line and band correspond, respectively, to the median value and 68% credible interval of ${Y}_{{\rm{SZ}}}^{{\rm{ALMA}}}$ estimated from the ALMA+ACA samples for an A10 UP pressure profile. The number of equal-mass subhalos with individual SZ fluxes equal to ${Y}_{{\rm{SZ}}}^{{\rm{ALMA}}}$ and required to obtain a total mass equal to the dynamical estimates is hence found to be $n \simeq 2.57$.

Extended Data Fig. 3 Impact of the number *n* of compact sources on the model reconstruction.

From top to bottom, denoting with ${\mathcal{Z}}$ the Bayesian evidence for each model: effective significance ${{\sigma }}_{{\rm{e}}{\rm{f}}{\rm{f}}}^{{\rm{x}}}={\rm{s}}{\rm{g}}{\rm{n}}({Z}_{n}^{{\rm{x}}}-{Z}_{0}^{{\rm{x}}})\cdot {(2|{Z}_{n}^{{\rm{x}}}-{Z}_{0}^{{\rm{x}}}|)}^{1/2}$ with respect to the null, data-only run ($n = 0$) of the models comprising only the radio-source component (x = pts) and the one including a SZ term (x = pts + sz); variation of the effective significance $\Delta {{\sigma }}_{{\rm{e}}{\rm{f}}{\rm{f}}}^{{\rm{x}}}={\rm{s}}{\rm{g}}{\rm{n}}$

$({Z}_{n}^{{\rm{x}}}-{Z}_{n-1}^{{\rm{x}}})\cdot {(2|{Z}_{n}^{{\rm{x}}}-{Z}_{n-1}^{{\rm{x}}}|)}^{1/2}$ in the radio-source-only and full models owing to the increment in the number of point-like components; effective significance ${{\sigma }}_{{\rm{e}}{\rm{f}}{\rm{f}}}^{{\rm{s}}{\rm{z}}}={\rm{s}}{\rm{g}}{\rm{n}}({Z}_{n}^{{\rm{p}}{\rm{t}}{\rm{s}}+{\rm{s}}{\rm{z}}}-{Z}_{n}^{{\rm{p}}{\rm{t}}{\rm{s}}})\cdot {(2|{Z}_{n}^{{\rm{p}}{\rm{t}}{\rm{s}}+{\rm{s}}{\rm{z}}}-{Z}_{n}^{{\rm{p}}{\rm{t}}{\rm{s}}}|)}^{1/2}$ of the SZ component with respect to the respective radio-only model; deviation in right ascension (ΔRA) and declination (ΔDec) from the coordinates of the Spiderweb galaxy (11 h 40 m 48.34 s, −26° 29′ 08.55″) as for Extended Data Table 2; mass parameter M_{500}; total flux of the radio model. After a swift evolution in the model evidence and parameters for $n \leq 5$ (grey area), the model nearly stabilizes around the reference solution (that is, for $n = 8$, marked as a black square in all panels) before incurring a degradation of its statistical significance at $n \geq 9$ (as clearly shown in the second panel by the negative $\Delta {\sigma }_{{\rm{eff}}}^{{\rm{pts}}}$ at $n = 9$). We finally note that, despite both being plotted, the terms ${\sigma }_{{\rm{eff}}}^{{\rm{pts+sz}}}$ and $\Delta {\sigma }_{{\rm{eff}}}^{{\rm{pts+sz}}}$ are not visible because, on the scales considered in the first two panels, they practically overlap with and are indistinguishable from the respective radio-source points ${\sigma }_{{\rm{eff}}}^{{\rm{pts}}}$ and $\Delta {\sigma }_{{\rm{eff}}}^{{\rm{pts}}}$. The error bars for each point denote the amplitude of the 68% (about 1σ) credible interval for the corresponding posterior distribution.

Source data

Extended Data Fig. 4 Variation of image-space residuals with respect to the number of point-like components used to model the extended radio emission.

Consistent with Extended Data Fig. 1, the images in all panels are generated considering only the data at *uv* distances smaller than 65 kλ. For the sake of simplicity, we did not apply any deconvolution procedure to the above

images. In fact, the large-scale patterns observed in the top-left panel ($n = 4$) reflect the non-uniform response pattern of the available visibility data.

Extended Data Fig. 5 Extended radio source from the sparse modelling.

Shown in the background is the maximum-a-posteriori model of the surface brightness for the radio source associated with the Spiderweb galaxy obtained from the sparse-image reconstruction discussed in Methods. The white contours denote the radio emission as measured from the VLA X-band[8,9] data (see also Fig. 3). To facilitate the comparison, we smoothed the sparse ALMA+ACA model to the same angular resolution as that of the VLA data. Except for a limited number of isolated and low-significance pixels, it is possible to observe a good agreement between the morphology of the reconstructed source and the VLA observation.

Extended Data Fig. 6 Diffuse gas around the Spiderweb galaxy.

The three panels show a comparison between the SZ effect (left) observed in the ALMA+ACA data with the X-ray emission measured by Chandra (centre) and the VLT/FORS1 image[19,30,127] of the extended Lyα nebula (right) within the Spiderweb complex. The ALMA+ACA SZ data are imaged using a natural weighting scheme after subtracting the sparse-modelling solution for the radio source. To reduce the dynamic range of the X-ray map, we suppress the strong emission from the central AGN by fitting and subtracting a point-like model to the data. The hatched region denotes the signal dominated by the inverse Compton X-ray emission associated with AGN or the radio jet[8,9,25], rather than the thermal bremsstrahlung from diffuse hot gas. For reference, we include in dark red the same VLA contours as in Extended Data Fig. 5. When discussing the extended X-ray signal from the ICM, we will then refer only to the region outside this mask (that is, excluding the radio jet). The contours in all the panels trace the decrement in the ALMA+ACA residual image after the application of a 20-kλ Gaussian *uv* taper to enhance the large-scale features associated with the SZ effect. The levels start at 2σ and increase by a factor of 0.5σ, with $\sigma = 5.30$ μJy beam^{-1}, as measured from the corresponding

jackknifed image. We mark with a white triangle the position of the Spiderweb galaxy. The ellipses in the lower-left corner of the left panel denote the beams for the ALMA+ACA SZ images at the original (solid ellipse) and *uv*-tapered (dotted ellipse) resolutions. Although the low significance and, in the case of the ALMA data, complex effects associated with interferometric filter hamper the possibility of observing an exact correspondence between the SZ and X-ray images, both of these manifest an asymmetry towards the northeast in their morphological signatures, suggesting an association of the two signals with a hot feature in the diffuse intracluster gas. On the other hand, the Lyα halo occupies the region corresponding to the southwest deficiency observed in both the SZ and the X-ray data. This result might imply that the very core of the Spiderweb galaxy is dominated by cold and warm gas that has not yet been shock-heated to the virial temperature.

Extended Data Fig. 7 Mock observations of simulated galaxy protoclusters.

Compared are the input Compton *y* maps from the Dianoga simulation set (left) and the images reconstructed from the simulated interferometric observations without (centre) and with (right) adding the jackknife-based noise realization. All maps are centred on the centroid of the dark-matter halo of each structure (also denoted as dotted crosshairs in the right panels), roughly corresponding (modulo a negligible displacement $\lesssim$6 kpc) to the position of the most massive galaxy in each system. The top row shows the maps for a simulated halo selected to generate a SZ signal with amplitude consistent with that measured for the Spiderweb protocluster. For comparison, we show in the middle row the most massive halo in the Dianoga sample. The bottom row corresponds instead to a simulated halo exhibiting a notable offset between the centre of the dark-matter halo and the peak in the pressure distribution (that is, SZ signal), with an overall displacement similar to what was found for the SZ footprint and the central galaxy in the Spiderweb complex. The asterisk on the colour scale label denotes the case in which the model surface brightness map is rescaled to match the signal-to-noise ratio of the ALMA+ACA measurements and, in turn, enhance the corresponding mock observation in the right panel. The white square in the left panels denotes the field of view of the mock

observation maps in the central and right columns. To facilitate the comparison, we set the colour map for the right frames equal to the one used in Extended Data Fig. 6. The solid contours in the right panels trace the noiseless mock observations in the central column, with contours starting from 1σ and increasing by 1σ steps (for $\sigma = 4.10$ μJy beam^{-1}). On the one hand, the limited signal-to-noise ratio for any of the halos of the sample highlights the need for deeper observations in view of a proper characterization of the morphological and physical properties of the forming ICM. On the other hand, the identification in the Dianoga set of halos with local asymmetries in their pressure distribution supports the case for the substantial offset observed in the ALMA+ACA data between the SZ signal and the Spiderweb galaxy not to be representative of large-scale displacements but the result of a non-trivial combination of irregular morphology and large-scale interferometric filtering.

Extended Data Table 1 Best-fit parameters for the model of the extended source

Extended Data Table 2 Comparison of the best-fit parameters for different SZ models

Supplementary information

Peer Review File

Source data

Source Data Fig. 1

Source Data Fig. 2

Source Data Extended Data Fig. 3

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Di Mascolo, L., Saro, A., Mroczkowski, T. *et al.* Forming intracluster gas in a galaxy protocluster at a redshift of 2.16. *Nature* **615**, 809–812 (2023). https://doi.org/10.1038/s41586-023-05761-x

- Received: 28 May 2022
- Accepted: 25 January 2023
- Published: 29 March 2023
- Issue Date: 30 March 2023
- DOI: https://doi.org/10.1038/s41586-023-05761-x

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

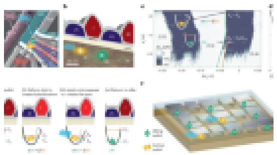
Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

Clues to the assembly of an infant galaxy cluster

Nature Research Briefing 29 Mar 2023

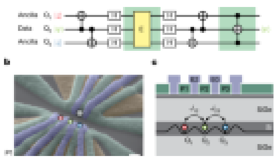
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05761-x



On-demand electrical control of spin qubits

12 January 2023

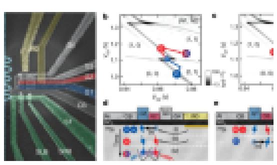
Will Gilbert, Tuomo Tanttu, … Andrew S. Dzurak



Quantum error correction with silicon spin qubits

24 August 2022

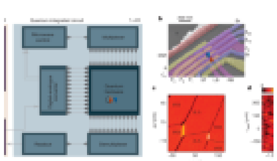
Kenta Takeda, Akito Noiri, … Seigo Tarucha



Fidelity benchmarks for two-qubit gates in silicon

13 May 2019

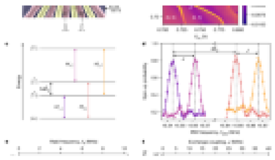
W. Huang, C. H. Yang, … A. S. Dzurak



Universal quantum logic in hot silicon qubits

15 April 2020

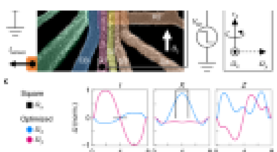
L. Petit, H. G. J. Eenink, … M. Veldhorst



[Fast universal quantum gate above the fault-tolerance threshold in silicon]

19 January 2022

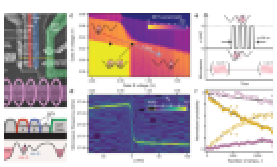
Akito Noiri, Kenta Takeda, … Seigo Tarucha



[Silicon qubit fidelities approaching incoherent noise limits via pulse engineering]

15 April 2019

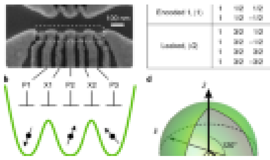
C. H. Yang, K. W. Chan, … A. S. Dzurak



[Coherent spin qubit transport in silicon]

05 July 2021

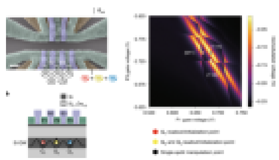
J. Yoneda, W. Huang, … A. S. Dzurak



[Quantifying error and leakage in an encoded Si/SiGe triple-dot qubit]

15 July 2019

Reed W. Andrews, Cody Jones, … Matthew G. Borselli



[Quantum tomography of an entangled three-qubit state in silicon]

07 June 2021

Kenta Takeda, Akito Noiri, … Seigo Tarucha

- Article
- Open Access
- Published: 06 February 2023

Universal logic with encoded spin qubits in silicon

- Aaron J. Weinstein ,
- Matthew D. Reed,
- Aaron M. Jones,
- Reed W. Andrews,
- David Barnes,

- Jacob Z. Blumoff,
- Larken E. Euliss,
- Kevin Eng,
- Bryan H. Fong,
- Sieu D. Ha,
- Daniel R. Hulbert,
- Clayton A. C. Jackson,
- Michael Jura,
- Tyler E. Keating,
- Joseph Kerckhoff,
- Andrey A. Kiselev,
- Justine Matten,
- Golam Sabbir,
- Aaron Smith,
- Jeffrey Wright,
- Matthew T. Rakher,
- Thaddeus D. Ladd &
- …
- Matthew G. Borselli

Show authors

Nature **volume 615**, pages 817–822 (2023)

- 6290 Accesses
- 4 Citations
- 76 Altmetric
- Metrics details

Abstract

Quantum computation features known examples of hardware acceleration for certain problems, but is challenging to realize because of its susceptibility to small errors from noise or imperfect control. The principles of fault tolerance may enable computational acceleration with imperfect

hardware, but they place strict requirements on the character and correlation of errors[1]. For many qubit technologies[2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21], some challenges to achieving fault tolerance can be traced to correlated errors arising from the need to control qubits by injecting microwave energy matching qubit resonances. Here we demonstrate an alternative approach to quantum computation that uses energy-degenerate encoded qubit states controlled by nearest-neighbour contact interactions that partially swap the spin states of electrons with those of their neighbours. Calibrated sequences of such partial swaps, implemented using only voltage pulses, allow universal quantum control while bypassing microwave-associated correlated error sources[1,22,23,24,25,26,27,28]. We use an array of six ^{28}Si/SiGe quantum dots, built using a platform that is capable of extending in two dimensions following processes used in conventional microelectronics[29]. We quantify the operational fidelity of universal control of two encoded qubits using interleaved randomized benchmarking[30], finding a fidelity of 96.3% ± 0.7% for encoded controlled NOT operations and 99.3% ± 0.5% for encoded SWAP. The quantum coherence offered by enriched silicon[5,6,7,8,9,16,18,20,22,27,29,31,32,33,34,35,36,37], the all-electrical and low-crosstalk-control of partial swap operations[1,22,23,24,25,26,27,28] and the configurable insensitivity of our encoding to certain error sources[28,33,34,38] all combine to offer a strong pathway towards scalable fault tolerance and computational advantage.

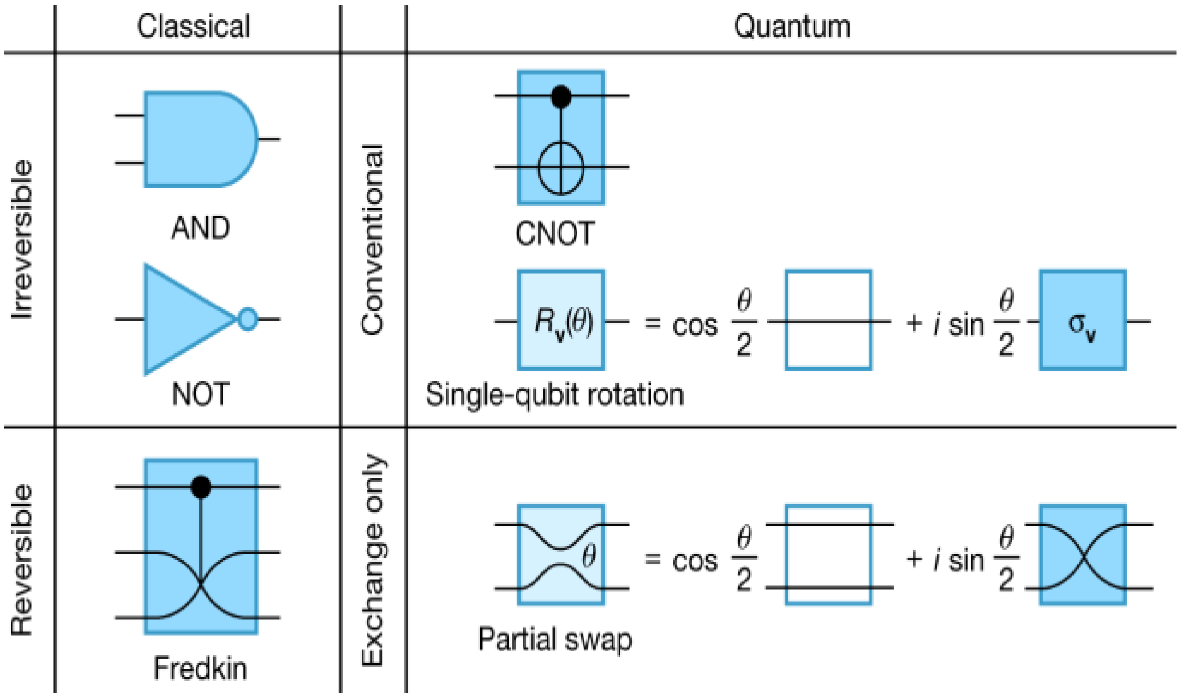
Main

The ultimate requirements for useful quantum hardware are set by fault tolerance (FT), whereby information is encoded in a way that contains and negates errors with a combination of redundancy, symmetry and careful scheduling of operations. FT requires, in part, that qubits be well-isolated from microscopic sources of noise and controlled with precision and high speed, all in a platform capable of scaling to sizes of computational relevance. Achieving the necessary scale favours lithographically defined qubits such as superconducting transmons[39] or single electron spins in Si quantum dots[22]. These approaches have enjoyed significant recent progress in scaling[3,40], control fidelity[5,6,7,41] and advanced fabrication[42,43]. Crucially,

however, FT also depends sensitively on the structure and correlation of the errors it is responsible for mitigating. Accordingly, qubit platforms that seem to satisfy the scale and fidelity requirements of quantum algorithms could nevertheless fail to achieve FT and so too fall short of computational acceleration. One significant source of possible correlated noise results from finite and varying energy splittings between qubit states, featured in most single-spin and transmon qubit systems. The phase evolution set by those splittings must be stable and continuously tracked, and conventional methods of control based on resonant driving between such splitting present challenges with respect to crosstalk.

Here we report on the realization of an alternative approach to quantum computation (QC) that uses degenerate qubit states and coherent partial swap interactions, the combination of which intrinsically avoids the FT challenges related to finite qubit energy splittings. Conventionally, qubits are controlled with arbitrary unitary rotations such as $R_{\mathbf{x}}(\theta) = \cos(\theta/2)\,\mathrm{id} + i \sin(\theta/2)\,\sigma_x$, where id is the identity and σ_x is the quantum analogue of the classical NOT, associated with adding or removing energy from a qubit. The combination of $R_{\mathbf{v}}(\theta)$ (for arbitrary axis **v** and angle θ) and a two-qubit gate such as controlled NOT (CNOT) constitutes a universal quantum gate set[44], analogous to the universal NOT and AND gates for classical bits (Fig. 1). This quantum gate set is able to traverse the very large continuous special unitary (SU) group of SU(2^n), where n is the number of qubits, rather than the (still very large) discrete symmetric group of 2^n bitstrings in the case of classical digital computers. Our alternative method of qubit control evokes the classical 'conservative logic'[45] of controlled-swapping of bits, which, as a Hamming-weight-conserving operation, operates on a smaller set of bitstrings scaling as $2^n/n$ for large n. This logarithmic overhead in bit number may be an acceptable cost for the benefit, in principle, of reducing heat associated with bit flips. In our quantum case, we apply arbitrary unitary superpositions of swapping and not swapping spins j and k of an n-spin array, that is, 'partial swaps', written as $U_{jk}(\theta) = \cos(\theta/2)\,\mathrm{ID} + i \sin(\theta/2)\,\mathrm{SWAP}_{jk}$ for any angle θ. Such a quantum computer could traverse SU($\approx 2^n/n^{3/2}$) for large n (ref. [23]).

Fig. 1: Example universal gate sets for four computational models.



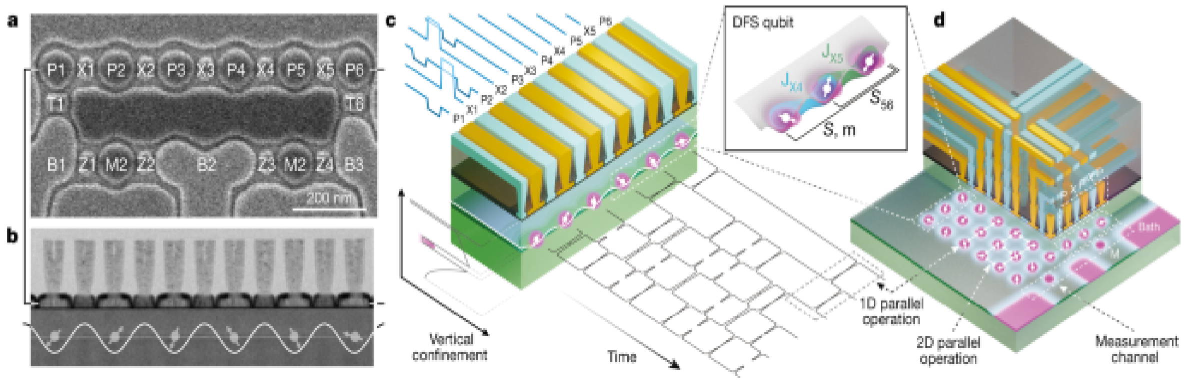
Top left, digital Boolean logic is universal with NOT and AND. Top right, single-spin-based quantum, with rotations about vector **v**, $R_{\mathbf{v}}(\theta)$ and CNOT. Bottom left, all Boolean logic may be performed reversibly within a subsystem using the Fredkin gate. Likewise, bottom right, all quantum logic may be performed within a subsystem using partial swaps.

Our demonstration uses a further refined subsystem, encoding $n/3$ qubit states in n spins. The remaining spin states outside this encoding are called leakage states, and their availability using controlled partial swapping provides a necessary resource for multi-qubit encoded universal quantum logic[1,24]. The linearly reduced qubit number and control-complexity overhead of our encoding relative to single-spin-based qubits enables the storage and control of qubit information in gapless quantum states. Indeed, this embedding was first theoretically proposed more than 20 years ago[1,23,24,25] precisely to avoid the correlated error whereby a global magnetic field fluctuation could dephase every qubit in a single-spin-based quantum computer at once. Such highly correlated errors would be ruinous to FT. By using only partial swaps in a global uniform field, this type of correlated error is avoided entirely; such encodings were hence called noiseless codes[25] or decoherence-free subsystems (DFSs)[1].

We implement the partial swap operation needed for DFS qubits with the calibrated activation of the contact exchange interaction between lithographically defined quantum dots. This highly local interaction, which arises as a consequence of the Pauli exclusion principle, is also commonly used for two-qubit operations in single-spin qubits and in double-spin singlet-triplet qubits[22] in conjunction with mechanisms to differentiate spin-splitting energies (for example, magnetic field gradients). For exchange-only (EO) operation, however, a critical design difference is the choice to remove as much variation in spin splittings as possible to maintain encoded-state degeneracy. EO operation has been demonstrated previously in GaAs[26] and Si-based[27,35] triple-quantum dots, but those encodings were limited to a single qubit with no ability to use leakage states to complete encoded universal quantum logic[1,23].

Our device confines electrons in a 5-nm-thick silicon well that is isotopically enhanced to a residual 800-ppm content of ^{29}Si; the well is surrounded by isotopically natural $Si_{0.3}Ge_{0.7}$ barriers. Isotopic enhancement reduces gradients in the effective magnetic field seen by electron spins because of hyperfine interactions[22], which constitute the dominant source of undesired spin-splitting variation and resulting control error for our encoding. Lithographically patterned metallic gates with tuned biases establish the electrostatic potentials that laterally confine and control interactions between electrons in the six quantum dots comprising the qubits and the two measurement dots (Fig. 2a,b). Voltages on plunger gates (P1–P6) accumulate a single electron in each dot and exchange gates (X1–X5) control the interaction strength between neighbouring electrons. Extra tunnel (T1, T6), bath (B1–B3) and charge sensor gates (M1, M2, Z1–Z4) (Fig. 2a) enable initialization and readout[46].

Fig. 2: A six-quantum-dot, two-qubit SLEDGE device in silicon.



a, Top-view scanning electron micrograph of the etched TiN single metal layer, defining the device’s plunger (P), exchange (X), tunnel (T), bath (B), barrier (Z) and measure-dot (M) electrode gates. Two DFS qubits are formed with the P1–P3 and P4–P6 dots and connected by a single exchange gate (X3). Unlabelled patches of metal are field gates, held at constant bias. **b**, Cross-sectional transmission electron micrograph of gate and TiN via electrodes, cut along the line of plunger and exchange gates in (**a**). Electrons depicted as circles with arrows are vertically confined by the Si/SiGe heterostructure boundaries and laterally confined by the induced electrostatic potential from the device gates. Al_2O_3/HfO_2 dielectrics appear as a black boundary around the electrode gates[29]. **c**, Qubit states are manipulated using sequences of voltage pulses (‘playing’ from the upper left in this schematic) that actuate nearest-neighbour exchange interactions. The interactions are principally modulated by X gates[33,35]. We perform entangling operations on the two-qubit encoded subspace with sequences of partial swaps such as the sequence shown playing to the lower right. **d**, Schematic of a future multilayer-BEOL SLEDGE device that enables scaling in the lateral dimension. Devices designed using this capability could feature several rails of coupled quantum dots and readout channels by leveraging the standard methods of BEOL signal interconnection used in the semiconductor microelectronics industry.

Our previous work[32,33,35,46,47,48], similar to other recent silicon qubit results[5,6,8,9,49], used an architecture that uses overlapping aluminium gates, which was challenging to reliably yield performant quantum dots at the level required for practical EO operation in extended arrays. Our present demonstration is instead based on the single-layer etch-defined gate electrode (SLEDGE) platform, recently detailed in Ha et al.[29], whose

improved yield results from separation in design and fabrication of the gate-defining layer (the 'single layer', which is a sheet of TiN-on-oxide above the silicon) from the back-end-of-line (BEOL) wiring, which contacts the gate layer through vertical via structures, as seen in Fig. 2. Although our device contains only two rows of quantum dots, the SLEDGE architecture readily allows extension of this geometry to two-dimensional (2D) arrays by stacking via-connected routing layers, as shown in Fig. 2d, following the BEOL interconnect strategies of modern silicon microprocessors[29]. Existing demonstrations of 2D arrays of semiconductor qubits use either carrier depletion in GaAs[10] or ring-like geometries of overlapping gates[11,12], both posing challenges to scaling that are circumvented by the via-based approach of SLEDGE. Arbitrarily large 2D arrays are still prevented by a finite number of feasible routing layers at the quantum dot pitch, but sufficient connectivity should nevertheless be available to enable 2D codes[13,14,50] when using the high-fidelity native encoded-SWAP operation that we demonstrate below.

We encode qubits using spins capable of partial swaps by associating quantum information with the total angular momentum quantum number of pairs of two spins, say 1 and 2, where states with $S_{12} = 0$ (the antisymmetric singlet state) map to $|0\rangle$ and states with $S_{12} = 1$ (the symmetric triplet state) map to $|1\rangle$. Hence S_{12} and S_{56} represent the two qubit states in our six-dot array (Fig. 2c). The addition of a third unpolarized and unmeasured spin to each qubit enables individual qubit control with only exchange, as shown in Kempe et al.[23], DiVincenzo et al.[24] and Andrews et al.[35], and elaborated in Methods. Unlike most other qubit systems[15,39], no additional physical mechanism is needed to generate entanglement between these encoded qubits. Sequences of exchange-based partial swaps must, however, be designed both to navigate leakage spaces and to be agnostic to the unknown relative states of the unpolarized spins (known as the gauge freedom), as derived by Fong and Wandzura (FW)[28,51] and further elaborated in Methods and Extended Data Figs. 5 and 6.

Qubit preparation and measurement is achieved with Pauli spin blockade. This is a common spin qubit technique that maps spin parity (symmetric versus antisymmetric configurations) to the more easily detected charge

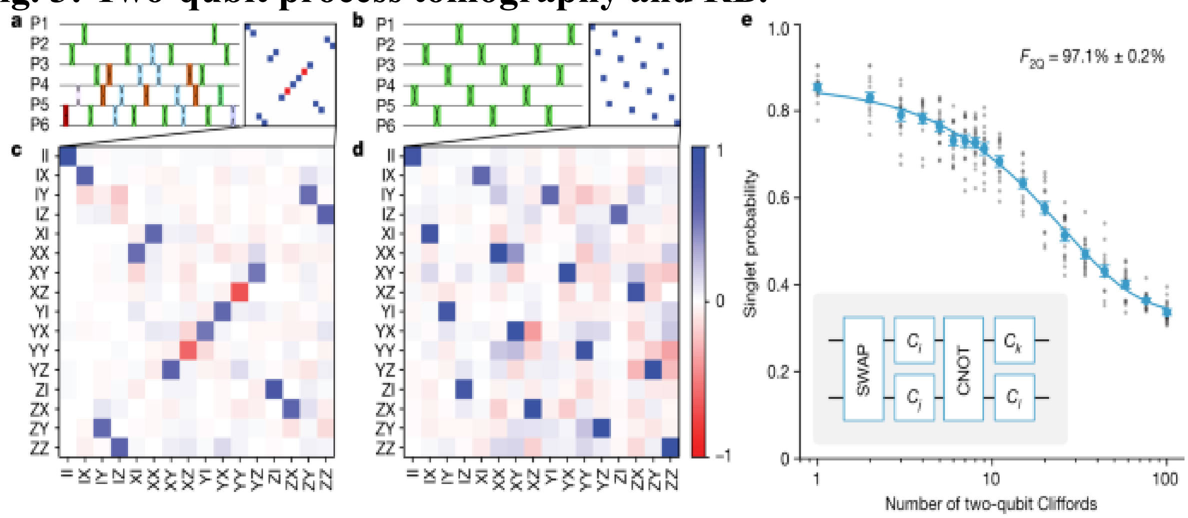
configuration[22]. We initialize the outermost dot pairs (P1/P2 and P5/P6), with the X3 gate connecting the innermost, uninitialized electron spins in dots P3 and P4. In Si/SiGe, the fidelity of these state preparation and measurement (SPAM) operations are, in part, limited by the valley excited-state energies of the dots, which we measure using photon-assisted tunnelling to be 70 μeV for dot P1 and 14 μeV for dot P6. The relatively poor valley splitting on P6 is consistent with observations of reduced SPAM fidelity when using sensor M2 to measure Pauli blockade on dots P5/P6, and is why we later prioritize P1/P2 measurement using sensor M1 when possible (Methods). M1-based measurement shows more than 96.0% ± 0.1% fidelity[46] in single-qubit characterizations (Extended Data Fig. 2). Valley splitting may be more reliably large in narrower wells[48], or possibly using other techniques;[52] SPAM fidelity was recently validated to reach more than 99.7% in a similar, narrower-well device[46].

We perform exchange-based control of our encoded qubits using interleaved sequences of baseband voltage pulses that activate nearest-neighbour interactions for a fixed duration t_{pulse}, with pulse-to-pulse spacing t_{idle}. A single exchange pulse is actuated by voltage-modulating gate Xn by approximately 100 mV (and neighbouring Pn and P($n+1$) gates by about −20 mV for capacitive compensation to achieve symmetric operation)[33]. This drives a partial swap $U_{n,n+1}(\theta)$ by temporarily raising exchange energy $J_{\mathrm{X}n}$ from a negligible value to $J_{\mathrm{X}n}/h \approx 100$ MHz (Fig. 2c). The specific angles θ we used for two-qubit gates are tabulated in Extended Data Table 1 and for single-qubit Clifford gates in Extended Data Table 2, although θ may be continuously tuned for arbitrary control. We provide control performance metrics in the Extended Data using methods described in Reed et al.[33] and Andrews et al.[35], including an evaluation of random magnetic field gradients due to residual nuclear spins in Extended Data Fig. 1a and control-signal crosstalk in Extended Data Fig. 2b. These metrics are explained further in Methods.

With encoding, SPAM and control in hand, we demonstrate two families of encoded two-qubit gates: CNOT (completing the universal gate set for quantum computing shown in Fig. 1) and SWAP. The latter is composed of 15 full spin swaps, and is the key resource for moving information through a

larger array of dots. We diagram the FW-CNOT and SWAP pulse sequences, decomposed as partial swaps, in Fig. 3a,b. The corresponding gate-specific voltages used for the FW-CNOT case are elaborated in Extended Data Fig. 3. As initial confirmation that we are performing the intended logical operations, we characterize the gates using quantum process tomography (QPT)[53], shown in Fig. 3c,d. We see qualitative agreement with expected tomograms, but, because of the weakness of our QPT inversion method to leakage and SPAM errors, and in particular the poor M2 SPAM in the present device, it is difficult to extract meaningful quantitative results from this protocol. This could, in principle, be mitigated with increased averaging and self-consistent gate set tomography[5,16,54], but we do not explore that here.

Fig. 3: Two-qubit process tomography and RB.



a,**b**, Exchange pulse diagrams and ideal process matrices of FW-CNOT (**a**) and SWAP (**b**) gates. The shading of exchange pulse boxes is proportional to the pulse exchange angle; detailed pulse angles may be found in Extended Data Tables 1 and 2. **c**,**d**, Maximum-likelihood estimates of measured quantum process matrices of FW-CNOT and SWAP gates. These data depend on joint measurement, so the relatively low M2 SPAM fidelity reduces contrast. The CNOT tomogram was measured on a similar device from the same wafer, with the same well width and gate pitch and comparable SPAM. **e**, Two-qubit RB with $t_{pulse} = 10$ ns, $t_{idle} = 5$ ns and applied magnetic field $B = 2.1$ mT. Individual dots represent the average from 500 shots of a single instance of a random Clifford sequence. Error

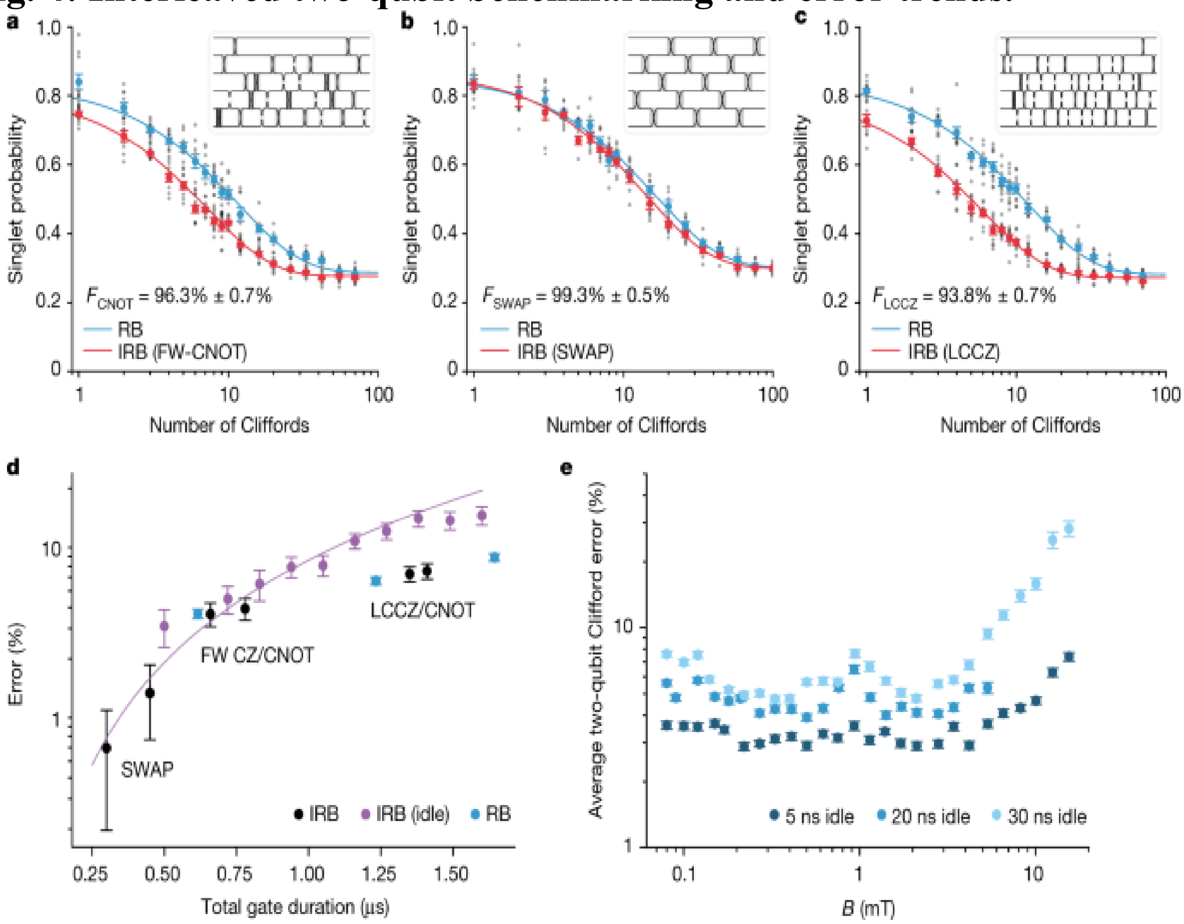
bars here represent the 1σ standard deviation calculated at each Clifford length. Two-qubit Clifford gates are compiled using FW-CNOT, SWAP and single-qubit Clifford gates C_x (inset).

Randomized benchmarking (RB) is our preferred method of characterizing gate performance as it is fast, simple, relatively insensitive to SPAM error and requires measurement of only one qubit[55,56]. In RB, a randomly selected sequence of gates that compile to the identity are chosen from a discrete group of qubit operations, typically the Clifford group. This choice depolarizes noise in the encoded subspace, allowing gate performance to be inferred by sweeping the sequence length and fitting return probability to an exponential decay. We generate two-qubit Cliffords by means of standard compilation rules[57] using the FW-CNOT entangling gate, the aforementioned SWAP gate and single-qubit Cliffords. Of the 11,520 two-qubit Clifford gates, 90% include a CNOT in our composition, 50% include the encoded SWAP and each has an average of 3.1 single-qubit Clifford gates; each operation thus contains 41.1 exchange pulses on average. As shown in Fig. 3e, we find an average two-qubit Clifford fidelity of 97.1% ± 0.2%, far better than suggested by the SPAM-afflicted QPT and comparable to contemporary single silicon spin qubit two-qubit RB errors[7,17], despite using a much larger number of exchange operations per Clifford. (This definition of fidelity incorporates errors due to leakage outside the encoded space but does not separate such leakage from encoded error as in the blind RB protocol[35], for which a two-qubit extension is an ongoing effort, see Methods and Extended Data Fig. 7.) The leakage calculation predicting the observed asymptote of $17/60 \approx 0.283$ is discussed in Methods.

We also measure the performance of individual gates of interest using interleaved randomized benchmarking (IRB)[30]. In this protocol, the operation in question is repeatedly interleaved between a random Clifford gate and the resulting decay is compared with a reference RB decay to infer the operation's fidelity. We perform IRB for FW-CNOT and encoded SWAP in Fig. 4a,b, finding a fidelity of 96.3% ± 0.7% and 99.3% ± 0.5%, respectively. The SWAP error is more than five times lower than that of CNOT, constituting a high-fidelity mechanism for moving data in constrained geometries using this technology. (Spin shuttling through empty dots[17,18,50,58] is a notable alternative.) A third gate we examine resolves a

new potential obstacle to FT posed by our encoding, namely, that gate error can now lead to qubit leakage and the FW-CNOT can spread that leakage throughout a register. To avoid this, a variant of FW-CNOT called the leakage-controlled controlled-*Z* (LCCZ) was designed to reduce the probability of leakage spreading to approximately the same probability as direct leakage. We find the LCCZ fidelity to be 93.8% ± 0.7% in this device; see Methods.

Fig. 4: Interleaved two-qubit benchmarking and error trends.



a–**c**, Benchmarking of FW-CNOT (**a**), SWAP (**b**) and LCCZ (**c**) at $B = 0$ mT, with inset exchange pulse diagrams. Here $t_{idle} = 20$ ns for FW-CNOT and LCCZ, but $t_{idle} = 10$ ns for SWAP; $t_{pulse} = 10$ ns for all. Note, the number of Cliffords excludes the final inverting Clifford. **d**, Error as a function of gate duration for selected gates (black) (Methods), for various interleaved idle periods (purple) and for two-qubit Cliffords (blue). Idle error increases with gate duration, closely following a theoretical estimate given by $(t_{idle}/T_2^*)^2$

(purple curve). Some gates show a significant deviation below this curve, indicating that they have some built-in magnetic noise insensitivity. **e**, Average two-qubit Clifford gate error as a function of *B* and pulse idle time. *B* is oriented in-plane and perpendicular to the dot array. We see consistent improvement for lower t_{idle}, eventually limited by the available bandwidth of the signal chain (not shown). As *B* increases, we first observe an improvement in fidelity above 200 μT, consistent with the suppression of transverse hyperfine magnetic gradients. The fidelity decreases for $B > 3$ mT because of induced paramagnetic gradients[32], see Extended Data Fig. 4. All dots in (**a**–**c**) represent the average from 1,000 shots of a single random Clifford sequence. All error bars in this figure correspond to 1σ standard deviation intervals.

The dominant error in our EO operation is due to residual hyperfine interactions from the 800 ppm of ^{29}Si nuclear spins and natural abundance of barrier ^{73}Ge spins. Evidence for this is provided by the scaling of the sequence error with duration (Fig. 4d), magnetic field dependence (Fig. 4e) and comparison to detailed quantitative numeric simulations, all elaborated in Methods. This varies from studies on single-spin qubits using micromagnetic field gradients, in which isotopic enhancement typically reduces magnetic noise beneath a floor set by charge noise[19,20,21]. EO is different, as it depends more critically on field uniformity, but is more resilient to charge noise, which we estimate as providing less than 6% of the observed error.

We have demonstrated encoded universal quantum logic operations, including single-qubit gates, CNOT, controlled-*Z* (CZ), LCCZ and SWAP, using only the exchange-based partial swap interaction and implemented in an extensible physical architecture. This control scheme is distinct from the traditional non-energy-conserving methods of other quantum technologies and may by itself offer near-term opportunities for analogue simulation[59,60]. In the longer term, although the EO DFS encoding offers fundamental advantages, achieving FT with it would require, in part, improved gate fidelity. Given the hyperfine contribution to our error, the primary pathway is to pursue faster control or improved isotopic purification. These refinements could be made independent of SPAM, fabrication yield and crosstalk performance, and are compatible with scaling to larger devices

using several back-end metal layers within the SLEDGE platform. This contrasts with lithographically defined microwave-actuated qubits, for which crosstalk reduction and site selectability typically require substantial device redesign with scale. An open question is whether our demonstrated coherent EO universal gate set may leverage states beyond our demonstrated encoding to enable other computational modalities, such as high-efficiency reversible classical computation[45] or permutational QC[61].

Methods

Encoded subspace notation

In the following sections, we elaborate sources of noise, the theoretical construction of two-qubit gates and further two-qubit benchmarking results. For these elaborations, a more detailed description of the encoded subspace is required, which is best described according to angular momentum quantum numbers. A single spin, say spin 1, has $S_1 = 1/2$ and projection $m_1 = \pm 1/2$. For two spins, labelled 1 and 2, the total angular momentum S_{12} can be $S_{12} = 0$, the singlet state for which $m_{12} = 0$, or $S_{12} = 1$, the triplet of states for which $m_{12} = -1$, 0 or 1. We notate these four states as $|S_{12}; m_{12}\rangle$. The smallest universally controllable EO qubit requires adding a third spin, for which the algebra of angular momentum provides the quantum number S_{123}, which can be 1/2 either by adding the single $S_3 = 1/2$ spin to the $S_{12} = 0$ singlet or by adding S_3 to the $S_{12} = 1$ triplet. These two choices for $S_{123} = 1/2$ comprise our encoded qubit, and the quadruplet of states with $S_{123} = 3/2$, which also results from adding S_3 to the $S_{12} = 1$ triplet, comprise our leaked states. We also have a total spin projection $m_{123} = -S_{123}, -S_{123} + 1, ..., S_{123}$. We thus notate these states here as $|S_{12}, S_{123}; m_{123}\rangle$ and our qubit states are $|0, 1/2; m_{123}\rangle$ and $|1, 1/2; m_{123}\rangle$. The m_{123} degree of freedom is called the gauge, and it is unaffected by exchange operation between any of the three spins. The Pauli spin blockade process on dots 1 and 2 provides initialization and measurement of only the S_{12} quantum number and has no impact on m_{123} (ref. [46]). Singlet measurements on spins 1 and 2 therefore measure the probability of $|0, 1/2; m_{123}\rangle$ independent of m_{123}, and provide no information

to distinguish encoded triplet $|1, 1/2; m_{123}\rangle$ states from leaked states $|1, 3/2; m_{123}\rangle$. More information about exchange-based gates in a single-qubit system is available in Andrews et al.[35].

Physical noise sources

We now discuss our techniques for characterizing two well-understood sources of noise and decoherence in our device, that is, magnetic noise and charge noise, and our techniques for limiting crosstalk error. This constitutes a more detailed discussion of Extended Data Figs. 1 and 2 than the main text could contain.

Magnetic or hyperfine noise

The DFS encoding we use is predicated on a homogeneous (although possibly time-varying) magnetic field from one dot to the next. Uniform magnetic fields are thus effectively ignored by the DFS encoding, whereas magnetic gradients drive relative precession of electron spin pairs, causing decoherence and leakage out of the encoded subspace. As such, magnetic gradients between spins cause error, both within the encoded subsystem and outside it (that is, leakage.) Magnetic gradient fluctuations, caused by the noisy magnetization of ^{29}Si and ^{73}Ge nuclear spins coupled to the electron spins by means of the contact hyperfine interaction, provide the largest contribution to our gate error[32].

We characterize the effective decoherence rate due to hyperfine noise, T_2*, by measuring the decay of a singlet prepared between two electrons and left in the low-exchange ‘idle’ configuration for a varying amount of time. The ensemble-averaged measurement projection decays to a value associated with the predicted mixture of encoded and non-encoded states: with probability 1/2 to the initial singlet $|0, 1/2; m_{123}\rangle$, 1/6 to the encoded triplet $|1, 1/2; m_{123}\rangle$ and 1/3 to the leaked state $|1, 3/2; m_{123}\rangle$[62]. Spin singlet pairs are initialized on either side of the device then shuttled to and from the desired location by consecutive π exchange pulses. Owing to the relatively poor SPAM fidelity on the M2 side of the device, spin pairs measured on the M2 side show deviations in the decay asymptote from the predicted value of

1/2. The 1/e point of the Gaussian decay defines T_2*, which we measured to be about 3.5 μs for six different spin pairs, as shown in Extended Data Fig. 1a. This timescale is further characterized in Kerckhoff et al.[32] and is fully expected for the present isotopic content. As expected for the number of hyperfine nuclei contributing to this dephasing, there is relatively little variation between each dot[32]. Other silicon qubits with varying levels of isotopic enhancement have been demonstrated elsewhere; for example Xue et al.[5] used similar 800-ppm Si and natural Ge, but Struck et al.[20] used 60-ppm ^{29}Si, and key demonstrations of long decoherence times approaching minutes have emerged from echo experiments from donor-bound electrons in 50-ppm ^{29}Si (ref. [31]). In single-dot experiments with thicker quantum wells and micromagnetic gradients, a key learning[19,20,21] is that isotopic purification of the silicon is sufficient to cause T_2* to be dominated by transduced charge noise, whereas in thinner quantum wells without gradients, such as the present device, nuclear hyperfine effects continue to dominate, with a clear pathway for improvement of increased isotopic enrichment[32,36].

Unlike charge noise, magnetic noise acts during both pulsing and idling[62] (for example, regardless of exchange energy), and so dephasing occurs continuously during evolution, leading to a total error scaling as $(t_{gate}/T_2^*)^2$. The prefactor of that scaling depends on how much the sequence permutes spins and decouples magnetic noise[32]. In Fig. 4d, we plot the IRB error of idle operations of increasing duration t. The calculated theoretical IRB error is $\varepsilon = (120/129)\ (t/T_2^*)^2$, in good agreement with the data. (The coefficient 120/129 is calculated by the convolution method presented in Merkel et al.[63].) For non-idle operations, error typically falls beneath the $\varepsilon \approx (t_{gate}/T_2^*)^2$ error level of ‘doing nothing’, because of the partial homogenization of gradient fields due to driven spin exchange. As an example, the LCCZ operation shown here has roughly half the infidelity of an equal-duration idle. The expected IRB error due to magnetic noise for a particular gate built from multipulse exchange sequences is evaluated by numeric simulation[35,62]. We find simulated magnetic error rates to be consistent with the measured data, although they are field dependent, and agreement is limited to uncertainties in the actual magnetic field witnessed at the quantum dot.

Although the EO encoding is explicitly immune to static and fluctuating global fields, global fields do affect the magnitude of local magnetic field gradients. In Fig. 4e, we plot average two-qubit Clifford fidelity measured using RB as a function of both applied field and t_{idle}. Here, the magnetic field is aligned in-plane with the device substrate and roughly perpendicular to the linear dot array. We find that error initially decreases by nearly a factor of two with increasing field strength, which is understood as the suppression of nuclear magnetic fluctuations transverse to the field direction, consistent with our numerical simulations[32].

At fields >3 mT, error increases again because of a combination of spin-orbit effects and Meissner screening effects from superconducting parts of the gate stack. The differences in Larmor frequency between the P1 dot and the P2–P5 dots are measured as in Eng et al.[27] and plotted in Extended Data Fig. 4, suggestive of a very small Meissner gradient of order $g|dB/dx|/h = 0.04$ (kHz/mT)/nm, but certainly also including a contribution from spin-orbit coupling[37,64]. The Meissner screening effect was much stronger in earlier devices using aluminium metal[32], but is much weaker in the present device because of the non-superconducting or weakly superconducting TiN gates used in the SLEDGE process[29]. The amount of spin-orbit (which depends on random Ge atom placement) and Meissner screening (which depends on deposition properties and geometry of TiN gates) vary significantly from dot to dot.

Charge or exchange noise

Another contribution to our gate error is charge noise, here manifest as exchange noise induced by fluctuations in the lateral trapping and tunnelling potentials. These fluctuations are due to either noise in the signal chain or defects in the gate stack and have a strong $1/f$ spectrum across many decades[65]. This type of noise induces error only during active exchange pulsing, as nearest-neighbour tunnel coupling is suppressed when the associated spins are idling. This idle error suppression results from the exponential scaling of the spin–spin exchange interaction with applied barrier potentials in general, and in particular from the large on/off ratios that are available with our tightly confining SLEDGE design[29]. As shown in

Extended Data Fig. 1b, we quantify this charge noise contribution to our error budget with an exchange oscillation Q-factor at a given J (typically $J/h \approx 100$ MHz). The product of J/h with the 1/e duration of the Gaussian decay envelope of exchange oscillations gives us a number of oscillations, N_{osc}. As a budgeting metric, N_{osc} is a superior parametrization, compared with the decay time, as the impact of charge noise on J scales as $|dJ/dV|$ for gate voltages V. If J were exactly exponential with voltage, N_{osc} would be independent of voltage, but it is not perfectly so as exchange is subexponential with X-gate voltage[33]. Similar to our magnetic noise heuristic, the estimated theoretical gate error from charge noise scales quadratically with this decay envelope, \(1/{N}_{{\rm{osc}}}^{2}\).

We repeatedly measure N_{osc} as a function of J along the symmetric operation axis, finding that N_{osc} reaches about 50 at 100 MHz, the frequency of operation for experimental convenience. Notably, and contrary to pure exponential activation, N_{osc} rapidly increases from <100 at $J/h = 1$ GHz to >1600 above $J/h \approx 20$ GHz, where we observe a sharp reduction in $|dJ/dV|$ as the potential barrier flattens (Extended Data Fig. 1c). This is the underlying reason that exchange is a subexponential function of voltage, such that the sensitivity to voltage fluctuations, which scales as $|\partial J/\partial V|$, reduces as J increases. We see in Extended Data Fig. 1 that, as exchange asymptotically approaches the approximate double-dot orbital energy, $|\partial J/\partial V|$ nearly vanishes, leading to a marked insensitivity to noise and thousands of exchange oscillations. This phenomenon has been previously reported in GaAs in a different operating regime[66]. The accessibility of this operation point in a SiGe accumulation mode device can be attributed to the large gate action of the SLEDGE design. To resolve coherent oscillations, which occur at exchange frequencies well above the 200-MHz Nyquist frequency of our arbitrary wave generators, we shift the arbitrary wave generator time basis, which is normally set to 2.5 ns, to within the range 2.5–5 ns. A continuous shift of the time basis within this range provides the smooth sampling needed to resolve coherent oscillations without aliasing, and thus to extract both the exchange rate, J/h, and N_{osc}. Although these exchange energies are too high for practical pulsed operation, they may prove valuable for microwave-sensitive EO encodings[67].

We perform an initial validation of our error model with single-qubit RB in Extended Data Fig. 2a. To properly account for leakage out of our encoded space, we use the 'blind' benchmarking technique described in Andrews et al.[35], whereby sequences are chosen to compile either to the identity or to one of the Pauli gates, and we analyse differences between the I, Z and X, Y branches. We find an average single-qubit Clifford error of $(1.1 \pm 0.1) \times 10^{-3}$ with a leakage error of $(3 \pm 1) \times 10^{-4}$. This is in approximate agreement with our simulated prediction of error from magnetic and charge noise alone, 5.0×10^{-4}, but these simulations do not include the effects of pulse distortion or other physical contributions such as Meissner screening or spin-orbit interactions. Evaluation of the impact of charge noise on two-qubit sequences is again done by numeric simulation, and it is found that, for these longer sequences, charge noise provides a smaller contribution to the error, <6%, as indicated in the main text.

Crosstalk

A low level of crosstalk is a key ingredient for practical scalability of a quantum register. Although the exchange interaction is intrinsically strongly limited to only nearest-neighbour action[22], thus limiting 'accidental' exchange for pairs not actively pulsed, control-signal crosstalk could still occur if attempting several exchange operations simultaneously in too small a region. The ultimate reason for this is the classical cross-capacitance of metal gates, which leads to a spurious signal in a neighbouring gate, an interaction which falls off as approximately $1/r^3$ for gates separated by physical distance r. The effects of this are mitigated in our system in three ways. First, because exchange is an approximately exponential function of its control voltage, a small spurious voltage signal would cause only an exponentially small deviation of exchange rate at 'idle'. This is because when spins are in the 'idle' configuration, exchange rates are very low ($J_{Xn}/h < 10$ kHz) and their derivatives relative to the small voltages that may be generated by cross-capacitance to actively controlled gates are also very low ($dJ/dV/h <$ kHz/mV).

By contrast, when activated by d.c. voltage pulses of order 100 mV, both exchange rates are fast ($J_{\mathrm{Xn}}/h \approx$ MHz to GHz) and derivatives large ($dJ/dV \approx$

J/(10 mV)), meaning that simultaneous operation of an exchange pulse for two nearby pairs in an array is best avoided. For the size of array used here, all spins are considered to be ‘nearby’ for the purposes of crosstalk. This is the reason for our second crosstalk mitigation, which is to ensure that only one spin pair is active at a time, and all others are idle. Future, larger arrays will rely on distant crosstalk signals being small enough to not cause unacceptable errors even with the large values of dJ/dV associated with active exchange, and therefore will allow simultaneous operation. In this case, the approximately $1/r^3$ scaling of voltage signal crosstalk will allow simultaneous operation as long as a zone of exclusion is respected; we expect this zone to be comparable in size to the present device. This constitutes only a relatively small overhead in time, which is well worth the cost of avoiding correlated errors from simultaneous control noise or the need for contextual calibration. We note that both of these ‘idle’ and ‘exchange’ regimes, associated respectively with our first and second crosstalk mitigation strategy, are a consequence of the high on/off ratio for exchange evident in Extended Data Fig. 1c and in Reed et al.[33].

Third, voltage pulsing is done using ‘virtual gates’[17,59] that compensate proximal barriers to keep nearest-neighbour electrons in idle and align the potential biases to symmetrize the exchange interaction to reduce the sensitivity of the system to charge noise[33].

To demonstrate this low susceptibility to crosstalk when operated with these mitigations in place, in Extended Data Fig. 2b we compare the IRB error of an idle period interleaved between Clifford gates with that of a control operation of equivalent duration performed on the neighbouring qubit. (For the reason explained in the discussion of the second mitigation strategy described above, we emphasize that exchange pulses are never simultaneously applied to generate this or any other data in this experiment.) As expected, because of the low voltage-crosstalk susceptibility at idle, the avoidance of simultaneous pulses and the use of symmetric operation (respectively mitigations one to three), there is no significant difference in error for the two cases to within the uncertainty of the RB measurement. This lack of degradation in performance from operations performed on a neighbouring qubit confirms that our control approach, whereby exchange

pulses are applied sequentially to the device without temporal overlap, enjoys minimal crosstalk error.

Theory of multi-qubit exchange-only pulse sequences

In this section we explain the structure of the entangling pulse sequences we demonstrated, including the purpose and construction of the LCCZ gate.

Unlike the single-encoded-qubit case, when considering operations on two encoded qubits on dots 1–6, the gauges of the two qubits, m_{123} and m_{456}, become important. Notating the total angular momentum of all six spins as S rather than S_{123456} and their projection as m instead of m_{123456} for brevity, the two $S_{123} = S_{456} = 1/2$ qubits may combine into an $S = 0$ subspace and an $S = 1$ subspace. Although exchange conserves m, still respecting gauge freedom, its action for interqubit operations does depend on S, which, in turn, is set by the relative value and phase of m_{123} and m_{456}. The gauge freedom, which can be safely ignored for single-qubit gates, must therefore be carefully considered with two-qubit operations.

In 2000, DiVincenzo et al.[24] provided a 19-pulse entangling gate sequence between two encoded qubits that required the total angular momentum of all six spins to be $S = 1$, which was imagined to be accomplished by polarizing the gauge of each qubit. However, the small spin-orbit effect and long spin relaxation times in silicon leave few good hardware choices for such polarization in our system, limiting the practicality of this approach. A decade later, the FW construction[28] was discovered by computational search using a genetic algorithm. This construction is a gauge-independent CNOT sequence, meaning it correctly performs the same CNOT sequence on the S_{12} and S_{45} degrees of freedom for both $S = 0$ and $S = 1$. Gauge independence allows each triple-dot EO qubit to be initialized as a pair of spin-singlet states and an unpolarized spin, as we do in the present demonstration.

In 2016, Zeuch and Bonesteel[51] showed that the FW sequence is in fact composed of three repetitions of a shorter primitive composite sequence acting on four spins, shown in Extended Data Fig. 5. This primitive sequence is a quasi-Fredkin (controlled-SWAP)[45] gate, swapping the gauge

m_{123} with m_4 only if $S_{12} = 1$, but not if $S_{12} = 0$. Applying this quasi-Fredkin gate to one qubit on spins 1, 2 and 3 and alternatingly on spins 4, 5 and then 4 again, a $S_{12} = 0$ condition will apply identity three times, whereas an $S_{12} = 1$ condition will swap spins 4 and 5, leaving m_{123} in its initial state (regardless of what that initial state is). These three uses provide a Fredkin gate with a three-spin EO qubit as control and two spins as target. If those two spins are the singlet–triplet pair of an EO qubit, this controlled-SWAP becomes an encoded CZ, the Fong–Wandzura controlled-*Z* (FWCZ), with no gauge dependence.

Compiling the primitive sequences together (for example, shuffling and combining commuting pulses, removing 2π pulses), one arrives at an entangling gate using 12 π/2 exchange pulses (that is, spin √SWAP gates) on just five fully connected spins. Assuring a CZ adds two more π/2 pulses, and adapting to the linear, nearest-neighbour coupled layout we use here with the measured singlet–triplet pairs on the ends of the array, the FWCZ ends up using all six spins with a further 12 π pulses (swaps on spins) for a total of 26 pulses. The sequence is shown in Extended Data Table 1. We may convert the CZ into CNOT by means of the construction shown in Extended Data Fig. 6a, adding two more pulses. To highlight the physical implementation of such sequences, we present an illustrative example in Extended Data Fig. 3 of the 28-pulse FW-CNOT sequence translated into experimentally accurate voltage waveforms required for device control. The sequences are compiled sequentially, so that no two pulses occur simultaneously, to avoid crosstalk (see Methods above).

The quasi-Fredkin gate on four spins discussed above has an undesired feature, however: if the EO qubit is in its $S_{123} = 3/2$ leaked state, then the gate applies a phase flip to the $S_{1234} = 2$ states relative to the $S_{1234} = 1$ states. Mathematically, no EO four-spin sequence can avoid this problem. As a result, when $S_{123} = 3/2$ and we apply this primitive operation as described for the FWCZ, then when $S_{12345} = 1/2$ and $S_{45} = 1$, the resulting unitary provides a phase flip, and when $S_{12345} = 3/2$, the resulting unitary applies a π rotation about an axis tipped an angle $\tan^{-1}[3\sqrt{(15/11)}]$ from the Bloch-sphere *z*-axis to the singlet–triplet qubit defined by S_{45}. These operations in general will leak the EO qubit including spins 4 and 5; even when applying

the FW gate perfectly, leakage will have spread from one qubit to the next. (Notably, leakage spreads only from the $|S_{12}, S_{123}; m_{123}\rangle$ qubit to the $|S_{45}, S_{456}; m_{456}\rangle$ qubit; if the $|S_{45}, S_{456}; m_{456}\rangle$ is leaked into $|1, 3/2; m_{456}\rangle$, then the gate behaves as intended for $S_{45} = 1$ regardless of S_{456}). This leakage spreading could be highly detrimental to FT, as we are unable to detect leakage directly in general and most quantum error correcting codes are ill-equipped to correct leakage even when it is detected.

The goal of the LCCZ gate is to avoid this leakage spreading when applying encoded CZ gates. The key insights to the LCCZ gate are that the unwanted leakage-induced phase flip or π rotation on the S_{45} qubit are both square roots of an identity operation and all single-qubit operations are identity on leakage spaces. Therefore, if we apply two FWCZ sequences with a single-qubit gate on the qubit composed of dots 1–3 in between, then we will have achieved some controlled-π-rotation gate on the encoded subspaces. The π-rotation angle depends on the choice of single-qubit operation and can be converted back to Pauli operators Z or X with single-qubit corrections. In particular, we use the construction shown in Extended Data Fig. 6b. The operator $\sqrt{Z}$ (called $S^{\dagger}$ in Extended Data Table 2, but not to be confused here with total spin) on $|S_{45}; m_{12}\rangle$ is simply another π/2 pulse on these spins, and the single-qubit rotations R, $R^{\dagger}$ and H may be readily derived as exchange sequences similar to those in Andrews et al.[35]. If the CZ in this construction is the FWCZ and the S_{12} qubit is leaky (that is, $S_{123} = 3/2$), then all single-qubit gates have no action and the two FWCZ gates combine into identity, leaving behind only the correctable single-qubit-encoded $\sqrt{Z}$ on the S_{45} pair (which is also $\sqrt{Z}$ on a $|S_{45}, S_{456}; m_{456}\rangle$ qubit). After some compiling, the resulting sequence is 46 pulses for a leakage-controlled CNOT and 44 pulses for an LCCZ; the compiled sequences are also shown in Extended Data Table 1.

The final sequence we demonstrate is encoded SWAP, which is not entangling, but is nonetheless a critical two-qubit gate for moving data and for RB. In many qubit modalities, the SWAP is more complex than the CNOT, as a typical construction uses three CNOT sequences. However, for EO qubits, SWAP is the one transversal operation of the underlying spins; for example, swapping all the underlying spins also constitutes a SWAP of

the encoded qubits. If our spins were fully connected, three π pulses to enact spin swaps would suffice. The 15-pulse sequence shown in Extended Data Table 1 performs the permutations needed to move spin information in linear, nearest-neighbour coupled architecture with the three-spin structure reflected about the centre. The timing of these spin-swap operations is chosen so that each spin spends approximately the same amount of time in each dot. A sequence that equalizes time exactly effectively decouples low-frequency magnetic noise;[34,38] the 15-pulse swap shown only partially completes such a permutative dynamical decoupling operation, but is nevertheless much more resilient to magnetic noise than other operations of similar duration.

Two-qubit benchmarking with bilateral readout and inversion rotations

Although all the RB results in the main text were generated with one-sided readout only, we present here benchmarking results with readout from both sides. As stated in the text, small valley splitting on dot P6 limits the SPAM fidelity and results in a meaningful reduction in measurement visibility. Nonetheless, as Extended Data Fig. 7 shows, the decay rates measured on either side are consistent; we observe similar gate performance using measurement on either side of the device.

In addition, we have explored the RB sequences with and without *X*-gate pre-rotations, reminiscent of the single-qubit blind RB protocol[35] and of character benchmarking on the larger space of encoded qubits[68]. Unlike the single-qubit case, in which linear combinations of the two measurement results can be used to differentiate error in the qubit computational space from leakage error, the two-qubit case navigates through a larger Hilbert space, which requires a greater number of measurements to distinguish those types of error in a similar manner. Although we have not yet formalized such a technique, we still emphasize that the decay rates on the singlet and triplet branches are consistent within the confidence intervals. We take this as a strong sign that the single exponential decay model for each channel is valid here, lending the possibility that linear combinations of exponential decays will enable more detailed leakage analysis in future efforts.

Finally, we explain the different asymptotic singlet probability values at large Clifford numbers for the different RB experiments we present. These may be understood by a relatively simple counting argument, which uses the fact that the total electron spin projection, m, is partially conserved in our experiments. This conservation is stronger at higher magnetic fields, where the mismatch of electron and nuclear Larmor frequencies suppress electron-nuclear flip flops and pulse sequences are insufficiently fast to drive the spin system to compensate. Although this conservation law is weaker at low magnetic fields, it is strong enough even in Earth's field that the following counting argument applies. For one-qubit RB, there are three states for three spins at constant value of $m_{123} = \pm 1/2$; in the $|S_{12}, S_{123}; m_{123}\rangle$ notation: $|0, 1/2; m_{123}\rangle$, $|1, 1/2; m_{123}\rangle$ and $|1, 3/2; m_{123}\rangle$. Magnetic noise will, under application of a large number of Clifford sequences, scramble these and result in a probability of 1/3 of measuring $S_{12} = 0$ at sequence end. For two-qubit RB, m is the projection across all six spins, and half of experiments begin with $m = 0$, a quarter begin with $m = 1$ and another quarter begin with $m = -1$, depending on the random gauge of the two initialized qubits. There are $6!/(3!3!) = 20$ ways to obtain $m = 0$ and $6!/(4!2!) = 15$ ways to obtain $m = \pm 1$. Six of the 20 $m = 0$ states have $S_{12} = 0$ and four of the 15 $m = 1$ states have $S_{12} = 0$, and so if these are again fully scrambled, the expected asymptote is $((6/20) + (4/15))/2 = 17/60 \approx 0.283$. (If we were to instead construct our Clifford gate set using the LCCZ gate, state mixing under m-conservation would be less complete and the theoretical asymptote would again become 1/3.) The asymptotic value observed when using the M2 readout is significantly higher than 1/3, as seen in Extended Data Fig. 7b, and is another indication of reduced SPAM performance on this side of the device.

Data availability

All data generated or analysed during this study are included in this published article and in its extended data.

References

1. Bacon, D., Kempe, J., Lidar, D. A. & Whaley, K. B. Universal fault-tolerant quantum computation on decoherence-free subspaces. *Phys. Rev. Lett.* **85**, 1758 (2000).

2. Blais, A., Huang, R. S., Wallraff, A., Girvin, S. M. & Schoelkopf, R. J. Cavity quantum electrodynamics for superconducting electrical circuits: an architecture for quantum computation. *Phys. Rev. A* **69**, 062320 (2004).

3. Arute, F. et al. Quantum supremacy using a programmable superconducting processor. *Nature* **574**, 505–510 (2019).

4. Xue, X. et al. Benchmarking gate fidelities in a Si/SiGe two-qubit device. *Phys. Rev. X* **9**, 021011 (2019).

5. Xue, X. et al. Quantum logic with spin qubits crossing the surface code threshold. *Nature* **601**, 343–347 (2022).

6. Mills, A. R. et al. Two-qubit silicon quantum processor with operation fidelity exceeding 99%. *Sci. Adv.* **8**, eabn5130 (2022).

7. Takeda, K., Noiri, A., Nakajima, T., Kobayashi, T. & Tarucha, S. Quantum error correction with silicon spin qubits. *Nature* **608**, 682–686 (2022).

8. Noiri, A. et al. Fast universal quantum gate above the fault-tolerance threshold in silicon. *Nature* **601**, 338–342 (2022).

9. Philips, S. G. J. et al. Universal control of a six-qubit quantum processor in silicon. *Nature* **609**, 919–924 (2022).

10. Mortemousque, P.-A. et al. Coherent control of individual electron spins in a two-dimensional quantum dot array. *Nat. Nanotechnol.* **16**, 296–301 (2021).

11. Ansaloni, F. et al. Single-electron operations in a foundry-fabricated array of quantum dots. *Nat. Commun.* **11**, 6399 (2020).

12. Hendrickx, N. W. et al. A four-qubit germanium quantum processor. *Nature* **591**, 580–585 (2021).

13. Vandersypen, L. M. K. et al. Interfacing spin qubits in quantum dots and donors—hot, dense, and coherent. *npj Quantum Inf.* **3**, 34 (2017).

14. Veldhorst, M., Eenink, H., Yang, C. H. & Dzurak, A. S. Silicon CMOS architecture for a spin-based quantum computer. *Nat. Commun.* **8**, 1766 (2017).

15. Monroe, C. et al. Programmable quantum simulations of spin systems with trapped ions. *Rev. Mod. Phys.* **93**, 025001 (2021).

16. Madzik, M. T. et al. Precision tomography of a three-qubit donor quantum processor in silicon. *Nature* **601**, 348 (2022).

17. Mills, A. et al. Shuttling a single charge across a one-dimensional array of silicon quantum dots. *Nat. Commun.* **10**, 1063 (2019).

18. Yoneda, J. et al. Coherent spin qubit transport in silicon. *Nat. Commun.* **12**, 4114 (2021).

19. Kawakami, E. et al. Gate fidelity and coherence of an electron spin in an Si/SiGe quantum dot with micromagnet. *PNAS* **113**, 11738–11743 (2016).

20. Struck, T. et al. Low-frequency spin qubit energy splitting noise in highly purified ^{28}Si/SiGe. *npj Quantum Inf.* **6**, 40 (2020).

21. Yoneda, J. et al. A quantum-dot spin qubit with coherence limited by charge noise and fidelity higher than 99.9%. *Nat. Nanotechnol.* **13**, 102 (2018).

22. Burkard, G., Ladd, T. D., Nichol, J. M., Pan, A. & Petta, J. R. Semiconductor spin qubits. *Rev. Mod. Phys.* (in the press); preprint available at https://arxiv.org/abs/2112.08863.

23. Kempe, J., Bacon, D., Lidar, D. A. & Whaley, K. B. Theory of decoherence-free fault-tolerant universal quantum computation. *Phys.*

Rev. A **63**, 042307 (2001).

24. DiVincenzo, D. P., Bacon, D., Kempe, J., Burkard, G. & Whaley, K. B. Universal quantum computation with the exchange interaction. *Nature* **408**, 339–342 (2000).

25. Zanardi, P. & Rasetti, M. Noiseless quantum codes. *Phys. Rev. Lett.* **79**, 3306–3309 (1997).

26. Medford, J. et al. Self-consistent measurement and state tomography of an exchange-only spin qubit. *Nat. Nanotechnol.* **8**, 654–659 (2013).

27. Eng, K. et al. Isotopically enhanced triple-quantum-dot qubit. *Sci. Adv.* **1**, e1500214 (2015).

28. Fong, B. H. & Wandzura, S. M. Universal quantum computation and leakage reduction in the 3-qubit decoherence free subsystem. *Quantum Info. Comput.* **11**, 1003–1018 (2011).

29. Ha, W. et al. A flexible design platform for Si/SiGe exchange-only qubits with low disorder. *Nano Lett.* **22**, 1443 (2021).

30. Magesan, E. et al. Efficient measurement of quantum gate error by interleaved randomized benchmarking. *Phys. Rev. Lett.* **109**, 080505 (2012).

31. Tyryshkin, A. M. et al. Electron spin coherence exceeding seconds in high-purity silicon. *Nat. Mater.* **11**, 143–147 (2012).

32. Kerckhoff, J. et al. Magnetic gradient fluctuations from quadrupolar ^{73}Ge in Si/SiGe exchange-only qubits. *PRX Quantum* **2**, 010347 (2021).

33. Reed, M. D. et al. Reduced sensitivity to charge noise in semiconductor spin qubits via symmetric operation. *Phys. Rev. Lett.* **116**, 110402 (2016).

34. Sun, B. et al. Full-permutation dynamical decoupling in triple-quantum-dot spin qubits. Preprint at https://arxiv.org/abs/2208.11784

(2022).

35. Andrews, R. et al. Quantifying error and leakage in an encoded Si/SiGe triple-dot qubit. *Nat. Nanotechnol.* **14**, 747 (2019).

36. Witzel, W. M., Carroll, M. S., Morello, A., Cywiński, Ł. & Sarma, S. D. Electron spin decoherence in isotope-enriched silicon. *Phys. Rev. Lett.* **105**, 187602 (2010).

37. Jock, R. M. et al. A silicon metal-oxide-semiconductor electron spin-orbit qubit. *Nat. Commun.* **9**, 1768 (2018).

38. West, J. R. & Fong, B. H. Exchange-only dynamical decoupling in the three-qubit decoherence free subsystem. *New J. Phys.* **14**, 083002 (2012).

39. Blais, A., Grimsmo, A. L., Girvin, S. M. & Wallraff, A. Circuit quantum electrodynamics. *Rev. Mod. Phys.* **93**, 025005 (2021).

40. Xue, X. et al. CMOS-based cryogenic control of silicon quantum circuits. *Nature* **593**, 205–210 (2021).

41. Sung, Y. et al. Realization of high-fidelity CZ and ZZ-free iSWAP gates with a tunable coupler. *Phys. Rev. X* **11**, 021058 (2021).

42. Bédécarrats, T. et al. A new FDSOI spin qubit platform with 40nm effective control pitch. In *2021 IEEE International Electron Devices Meeting (IEDM)* 1–4 (IEEE, 2021).

43. Zwerver, A. M. J. et al. Qubits made by advanced semiconductor manufacturing. *Nat. Electron.* **5**, 184–190 (2022).

44. Barenco, A. et al. Elementary gates for quantum computation. *Phys. Rev. A* **52**, 3457–3467 (1995).

45. Fredkin, E. & Toffoli, T. Conservative logic. *Int. J. Theor. Phys.* **21**, 219–253 (1982).

46. Blumoff, J. Z. et al. Fast and high-fidelity state preparation and measurement in triple-quantum-dot spin qubits. *PRX Quantum* **3**, 010352 (2022).

47. Borselli, M. G. et al. Undoped accumulation-mode Si/SiGe quantum dots. *Nanotechnology* **26**, 375202 (2015).

48. Chen, E. H. et al. Detuning axis pulsed spectroscopy of valley-orbital states in Si/SiGe quantum dots. *Phys. Rev. Appl.* **15**, 044033 (2021).

49. Corrigan, J. et al. Coherent control and spectroscopy of a semiconductor quantum dot wigner molecule. *Phys. Rev. Lett.* **127**, 127701 (2021).

50. Boter, J. M. et al. Spiderweb array: a sparse spin-qubit array. *Phys. Rev. Appl.* **18**, 024053 (2022).

51. Zeuch, D. & Bonesteel, N. E. Simple derivation of the Fong–Wandzura pulse sequence. *Phys. Rev. A* **93**, 010303 (2016).

52. McJunkin, T. et al. SiGe quantum wells with oscillating Ge concentrations for quantum dot qubits. *Nat. Commun.* **13**, 7777 (2022).

53. Merkel, S. T. et al. Self-consistent quantum process tomography. *Phys. Rev. A* **87**, 062119 (2013).

54. Nielsen, E. et al. Probing quantum processor performance with pyGSTi. *Quantum Sci. Technol.* **5**, 044002 (2020).

55. Knill, E. et al. Randomized benchmarking of quantum gates. *Phys. Rev. A* **77**, 012307 (2008).

56. Emerson, J., Alicki, R. & Zyczkowski, K. Scalable noise estimation with random unitary operators. *J. Opt. B, Quantum Semiclass. Opt.* **7**, 347–352 (2005).

57. Koenig, R. & Smolin, J. A. How to efficiently select an arbitrary Clifford group element. *J. Math. Phys.* **55**, 122202 (2014).

58. Seidler, I. et al. Conveyor-mode single-electron shuttling in Si/SiGe for a scalable quantum computing architecture. *npj Quantum Inf.* **8**, 100 (2022).

59. Hensgens, T. et al. Quantum simulation of a Fermi–Hubbard model using a semiconductor quantum dot array. *Nature* **548**, 70 (2017).

60. Wang, C. A. et al. Probing resonating valence bonds on a programmable germanium quantum simulator. Preprint at https://arxiv.org/abs/2208.11505 (2021).

61. Jordan, S. P. Permutational quantum computing. *Quantum Info. Comput.* **10**, 470–497 (2010).

62. Ladd, T. D. Hyperfine-induced decay in triple quantum dots. *Phys. Rev. B* **86**, 125408 (2012).

63. Merkel, S. T., Pritchett, E. J. & Fong, B. H. Randomized benchmarking as convolution: Fourier analysis of gate dependent errors. *Quantum* **5**, 581 (2021).

64. Tanttu, T. et al. Controlling spin-orbit interactions in silicon quantum dots using magnetic field direction. *Phys. Rev. X* **9**, 021028 (2019).

65. Connors, E. J., Nelson, J., Edge, L. F. & Nichol, J. M. Charge-noise spectroscopy of Si/SiGe quantum dots via dynamically-decoupled exchange oscillations. *Nat. Commun.* **13**, 940 (2022).

66. Dial, O. E. et al. Charge noise spectroscopy using coherent exchange oscillations in a singlet-triplet qubit. *Phys. Rev. Lett.* **110**, 146804 (2013).

67. Pan, A. et al. Resonant exchange operation in triple-quantum-dot qubits for spin–photon transduction. *Quantum Sci. Technol.* **5**, 034005 (2020).

68. Helsen, J., Roth, I., Onorati, E., Werner, A. & Eisert, J. General framework for randomized benchmarking. *PRX Quantum* **3**, 020357 (2022).

Acknowledgements

We thank J. B. Carpenter for assistance with all figures, and acknowledge significant technical contributions from E. Acuna, E. Chen, L. Edge, M. Gyure, A. Hunter, C. Jones, B. Maune, S. Merkel, I. Milosavljevic, E. Pritchett, R. Ross, A. Schmitz, C. Schnaible, B. Sun, R. Velunta, S. Wandzura and P. Williams.

Author information

Authors and Affiliations

1. HRL Laboratories, LLC, Malibu, CA, USA

   Aaron J. Weinstein, Matthew D. Reed, Aaron M. Jones, Reed W. Andrews, David Barnes, Jacob Z. Blumoff, Larken E. Euliss, Kevin Eng, Bryan H. Fong, Sieu D. Ha, Daniel R. Hulbert, Clayton A. C. Jackson, Michael Jura, Tyler E. Keating, Joseph Kerckhoff, Andrey A. Kiselev, Justine Matten, Golam Sabbir, Aaron Smith, Jeffrey Wright, Matthew T. Rakher, Thaddeus D. Ladd & Matthew G. Borselli

Contributions

M.T.R., T.D.L. and M.G.B. conceptualized and supervised the project. L.E.E., S.D.H., C.A.J., M.J., G.S., J.W., M.T.R. and M.G.B. designed and fabricated the devices. A.J.W., M.D.R., A.M.J., R.W.A., D.B., J.Z.B., K.E., D.R.H. and J.K. developed hardware, performed measurements and analysed data. J.M. and A.S. developed software to support measurements. B.H.F., T.E.K., A.A.K. and T.D.L. developed gate sequences, device models and simulations. A.J.W., M.D.R. and T.D.L. wrote the manuscript, with input and edits from all authors.

Corresponding authors

Correspondence to Aaron J. Weinstein or Thaddeus D. Ladd.

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks Michel Pioro-Ladriere and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Device performance metrics.

a, Magnetic dephasing of a spin singlet at the idle position (where $J \approx 0$), prepared on different dot pairs. The 1/e point of the Gaussian decay envelope defines $t = T_2^*$ and is a simple metric for characterizing the impact of substrate nuclear magnetic noise on qubit performance. We plot a $T_2^* =$ 3.5 μs envelope as a visual guide in dashed black. Most pairs are prepared and measured with high fidelity on the M1 side (left vertical axis); the P5/P6 pair uses M2 (right vertical axis) and has lower contrast. **b**, Charge noise impact on exchange oscillations for each exchange axis. This measurement is analogous to the one for T_2^* except measured at $J/h \approx 100$ MHz such that fluctuations in the exchange energy due to charge noise are the dominant source of decoherence. We parameterize the 1/e Gaussian decay point in

terms of the number of coherent oscillations, N_{osc}, that occur in that time. Each successive curve is offset on the y-axis by 0.5 and on the x-axis by 10 ns. Here, N_{osc} = 57.6, 49.2, 33.4, 70.9 and 42.3 at J/h = 119 MHz, 84.2 MHz, 114 MHz, 134 MHz and 103 MHz for axes J_{X1}–J_{X5}, respectively. **c**, N_{osc} as a function of J. Owing to the subexponential behaviour of exchange along the symmetric axis[33] (inset), we observe that N_{osc} increases with J. At $J/h \geq 17$ GHz, exchange asymptotes as a result of the flattening of the tunnel barrier to a value related to the double-dot orbital energy[66]. In this limit, where voltage excursions on numerous gates exceed ±200 mV, dJ/dV decreases significantly and N_{osc} rapidly increases to an observed maximum >1,600.

Extended Data Fig. 2 Single-qubit performance and inter-qubit crosstalk data.

a, Single-qubit performance for the P1, P2, P3 qubit with M1 readout. The 'blind' RB measurement[35] yields an average single qubit gate error of $(1.1 \pm 0.1) \times 10^{-3}$ and leakage error of $(3 \pm 1) \times 10^{-4}$. Using the methods of Blumoff et al.[46], we infer a SPAM fidelity on the M1 side from these data of 96.0% ± 0.1%. See Methods for more detailed discussion of this figure. **b**, Crosstalk characterization. IRB error on qubit Q_0 (using exchange energies J_{X1} and J_{X2}) while interleaving between each Clifford either periods of idle (blue) or composite Clifford gates (orange) on qubit Q_1 (using exchange energies J_{X5} and J_{X4}). The Q_1 composite sequences used here are X, XX, XXX, Y, YY, YYY, $S^\dagger$, $S^\dagger S^\dagger$ and $S^\dagger S^\dagger S^\dagger$. These gates, respectively, take 90 ns, 180 ns, 270 ns, 120 ns, 240 ns, 360 ns, 30 ns, 60 ns and 90 ns. The x-axis is the duration of the interleaved idle period or total duration of the composite pulse sequence on qubit Q_1. The black dashed line indicates modelled RB error due to hyperfine dephasing at each idle duration in a magnetic field of 20 μT as generated by a numeric simulation. As discussed in Methods, these data serve as a validation of our approach for limiting crosstalk error, which includes strict avoidance of simultaneous exchange pulses and low susceptibility to spurious signals in the device 'idle' configuration. All error bars in this figure correspond to 1σ standard deviation intervals.

Extended Data Fig. 3 A Fong–Wandzura CNOT and its corresponding voltage control waveform.

a,**b**, Pulse angles (**a**) defined on exchange axes are translated into rectangular voltage pulses (**b**) specified at the gates, with varying amplitudes along the corresponding symmetric axes. The voltage waveforms shown here are the exact voltage waveforms used in the RB experiment of Fig. 3c.

Extended Data Fig. 4 Paramagnetic gradients.

Magnetic gradients as a function of gate-to-gate separation in the lateral plane. Gradients are derived from singlet–triplet oscillations of a decay measurement, measured over various spatial separations, as described in Eng et al.[27]. A linear slope of 0.04 (kHz/mT) nm^{-1} is plotted here as a guide to the eye.

Extended Data Fig. 5 Mathematical construction of the FWCZ.

A primitive subsequence coupling all pairs of four spins with π/2 pulses (spin √SWAP) gates is shown in the dashed box, which, in the $S_{123} = 1/2$ (encoded) subsystem, is controlled by the S_{12} quantum number of the $|S_{12}, S_{123}; m_{123}\rangle$ qubit. The primitive is identity for $S_{12} = 0$ and swaps m_{123} with a fourth spin if $S_{12} = 1$. Three uses with alternating choice of the fourth spin completes a controlled-Z between singlet–triplet subsystems on S_{12} and S_{45}. If $S_{123} = 3/2$ (leakage subsystem), the gate applies an S_{12345}-dependent unitary U to S_{45}, with $U^2 = 1$. A full-pulse construction of this FWCZ including the extra π pulses (spin swaps) to adapt to a linear nearest-neighbour layout is shown in Extended Data Table 1.

Extended Data Fig. 6 LCCZ construction.

a, FW-CNOT is made from FWCZ (Extended Data Fig. 5) in the standard way: two Hadamard gates, notated H, as in Andrews et al.[35] and some compiling. **b**, The LCCZ sequence is made from two FWCZ sequences interspersed with single-qubit gates, and some compiling. Here $R = (I + iZ)/$

√2 for identity I and Pauli Z. Full-pulse constructions are shown in Extended Data Table 1.

Extended Data Fig. 7 Two-qubit RB with two-sided readout.

a,**b**, Benchmarking measurements on the M1 side (**a**) and M2 sides (**b**), with and without X-gate pre-rotations. Cliffords are compiled with the FW-CNOT entangling gate. We observe worse performance here than for the RB datasets in the main text as we operate here with $t_{\text{idle}} = 20$ ns, compared with $t_{\text{idle}} = 5$ ns in Fig. 3c, but performance here is consistent with that observed in Fig. 4e.

Extended Data Table 1 Two-qubit gate sequences
Extended Data Table 2 Single-qubit Clifford gate sequence

Supplementary information

Peer Review File

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Weinstein, A.J., Reed, M.D., Jones, A.M. *et al.* Universal logic with encoded spin qubits in silicon. *Nature* **615**, 817–822 (2023). https://doi.org/10.1038/s41586-023-05777-3

- Received: 01 March 2022
- Accepted: 31 January 2023
- Published: 06 February 2023
- Issue Date: 30 March 2023
- DOI: https://doi.org/10.1038/s41586-023-05777-3

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

This article is cited by

- **Noisy intermediate-scale quantum computers**
  - Bin Cheng
  - Xiu-Hao Deng
  - Dapeng Yu

Frontiers of Physics (2023)

- **[Universal control of a six-qubit quantum processor in silicon]**
  - Stephan G. J. Philips
  - Mateusz T. Mądzik
  - Lieven M. K. Vandersypen

  Nature (2022)

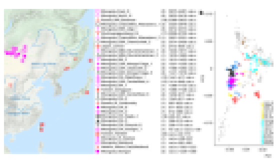
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05777-3



Searching for the roots of the first free African American community

26 November 2020

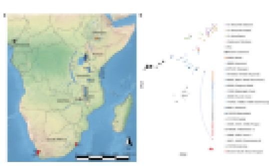
Beatriz Martínez, Filipa Simão, … Leonor Gusmão



Genomic insights into the formation of human populations in East Asia

22 February 2021

Chuan-Chao Wang, Hui-Yuan Yeh, … David Reich



Ancient DNA and deep population structure in sub-Saharan African foragers

23 February 2022

Mark Lipson, Elizabeth A. Sawchuk, … Mary E. Prendergast



[Ancient DNA reveals admixture history and endogamy in the prehistoric Aegean]

16 January 2023

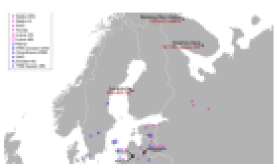
Eirini Skourtanioti, Harald Ringbauer, … Philipp W. Stockhammer



[The population history of northeastern Siberia since the Pleistocene]

05 June 2019

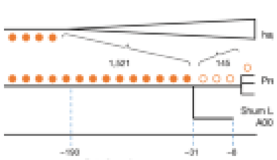
Martin Sikora, Vladimir V. Pitulko, … Eske Willerslev



[Ancient Fennoscandian genomes reveal origin and spread of Siberian ancestry in Europe]

27 November 2018

Thiseas C. Lamnidis, Kerttu Majander, … Stephan Schiffels



Ancient West African foragers in the context of African population history

22 January 2020

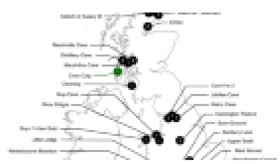
Mark Lipson, Isabelle Ribot, … David Reich



Palaeo-Eskimo genetic ancestry and the peopling of Chukotka and North America

05 June 2019

Pavel Flegontov, N. Ezgi Altınışık, … Stephan Schiffels



Ancient genomes indicate population replacement in Early Neolithic Britain

15 April 2019

Selina Brace, Yoan Diekmann, … Ian Barnes

- Article
- Open Access
- Published: 29 March 2023

Entwined African and Asian genetic roots of medieval peoples of the Swahili coast

- Esther S. Brielle ,
- Jeffrey Fleisher ,
- Stephanie Wynne-Jones ,
- Kendra Sirak,
- Nasreen Broomandkhoshbacht,
- Kim Callan,
- Elizabeth Curtis,
- Lora Iliev,
- Ann Marie Lawson,
- Jonas Oppenheimer,
- Lijun Qiu,
- Kristin Stewardson,
- J. Noah Workman,
- Fatma Zalzala,
- George Ayodo,
- Agness O. Gidna,
- Angela Kabiru,
- Amandus Kwekason,
- Audax Z. P. Mabulla,
- Fredrick K. Manthi,
- Emmanuel Ndiema,
- Christine Ogola,
- Elizabeth Sawchuk,
- Lihadh Al-Gazali,
- Bassam R. Ali,
- Salma Ben-Salem,
- Thierry Letellier,
- Denis Pierron,
- Chantal Radimilahy,
- Jean-Aimé Rakotoarisoa,

- Ryan L. Raaum,
- Brendan J. Culleton,
- Swapan Mallick,
- Nadin Rohland,
- Nick Patterson,
- Mohammed Ali Mwenje,
- Khalfan Bini Ahmed,
- Mohamed Mchulla Mohamed,
- Sloan R. Williams,
- Janet Monge,
- Sibel Kusimba,
- Mary E. Prendergast,
- David Reich &
- …
- Chapurukha M. Kusimba

Show authors

Nature **volume 615**, pages 866–873 (2023)

- 6841 Accesses

- 614 Altmetric

- Metrics details

Abstract

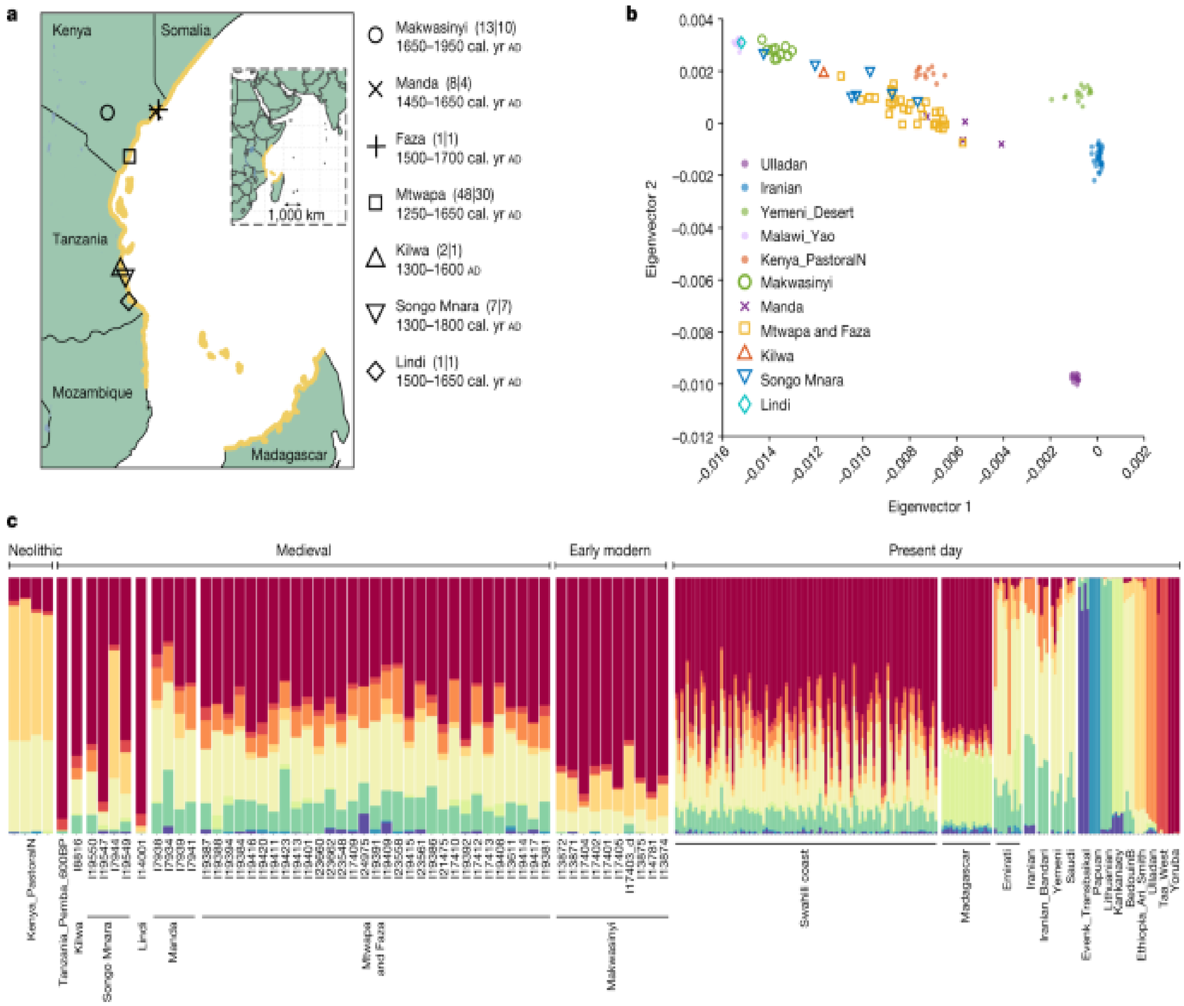
The urban peoples of the Swahili coast traded across eastern Africa and the Indian Ocean and were among the first practitioners of Islam among sub-Saharan people[1,2]. The extent to which these early interactions between Africans and non-Africans were accompanied by genetic exchange remains unknown. Here we report ancient DNA data for 80 individuals from 6 medieval and early modern (ad 1250–1800) coastal towns and an inland town after ad 1650. More than half of the DNA of many of the individuals from coastal towns originates from primarily female ancestors from Africa, with a large proportion—and occasionally more than half—of the DNA

coming from Asian ancestors. The Asian ancestry includes components associated with Persia and India, with 80–90% of the Asian DNA originating from Persian men. Peoples of African and Asian origins began to mix by about ad 1000, coinciding with the large-scale adoption of Islam. Before about ad 1500, the Southwest Asian ancestry was mainly Persian-related, consistent with the narrative of the Kilwa Chronicle, the oldest history told by people of the Swahili coast[3]. After this time, the sources of DNA became increasingly Arabian, consistent with evidence of growing interactions with southern Arabia[4]. Subsequent interactions with Asian and African people further changed the ancestry of present-day people of the Swahili coast in relation to the medieval individuals whose DNA we sequenced.

Main

The medieval and early modern Swahili culture of eastern Africa from the seventh century ad was defined by a set of shared features: a common language of African origin (Kiswahili), a shared predominant religion (Islam) and a geographic distribution in coastal towns and villages. People of the Swahili culture lived across a vast coastal region that included northern Mozambique, southern Somalia, Madagascar and the archipelagos of Comoros, Kilwa, Mafia, Zanzibar and Lamu[1] (yellow outlines in Fig. 1a). Millions of present-day coastal people identify as Swahili, although for many this is a secondary identity, with primary identities often being more based on town of origin, family history or traditional social status[5]. How people who identify as Swahili in the present day relate to people of the medieval and early modern Swahili culture has been difficult to elucidate in the absence of ancient DNA.

Fig. 1: Dataset overview.



a, Coastal areas associated with the medieval Swahili culture are shown in yellow. Sites represented in the ancient DNA samples are marked with black shapes. Numbers in parentheses are formatted *X*|*Y*, where *X* is the number of individuals for whom there are data, and *Y* is the number of individuals for whom we report high-resolution analyses. The chronology is given as the union of 95% confidence intervals for direct radiocarbon dates on the skeletons rounding to the nearest 50 years and is shown as calibrated (cal.) yr ad, or as ad for sites with only archaeological context (Supplementary Information; Extended Data Table 1). The base map was made with the Natural Earth R package using a CC BY license. Bodies of water were added from the RCMRD Geoportal with a CC BY license in R. All points were added and modifications were performed in R and Adobe Illustrator. The map is published for the first time in this Article. **b**, In a principal component analysis, eigenvector 1 correlates to variation maximized in sub-Saharan Africa, and eigenvector 2 correlates to Eurasian variation. **c**,

Ancestry component assignment using ADMIXTURE with $K = 9$ clusters (selected on the basis of low cross-validation errors, high log-likelihood scores, and a low number of reference populations to not overfit; groups that maximize each components are shown on the right). Individuals with sufficient data for high-resolution analysis are plotted in approximate chronological order from left to right. Ancient individuals are labelled and plotted at four times the width of present-day individuals.

The medieval largely autonomous towns and polities known as the Swahili states arose out of fishing and agropastoral settlements on the eastern African coast during the late first millennium ad[6]. First millennium ad sites on the littoral, beginning in the seventh century, were part of a shared material culture and practice network across the eastern African region[7]. These sites were engaged in the Indian Ocean trading system, facilitated by southwest monsoons from May to October that enabled merchant vessels to travel from India or the Arabian Peninsula to the eastern African coast, and northeast monsoons from November to March enabling their return in the same year[8].

Muslims were present from the eighth century ad, probably as a minority[9]. A major archaeological transition is evident during the eleventh century, with the establishment of new settlements and the elaboration of older ones with coral-built mosques and tombs, a set of changes generally understood as coinciding with the widespread adoption of Islam[10]. At this time, clearer distinctions also emerged between coastal ceramics and material traditions and those of inland assemblages[2,11], even as many aspects of material culture remained deeply linked with inland African groups.

The political and administrative independence of Swahili towns diminished in the sixteenth century as the Portuguese naval and economic dominance in the Indian Ocean spread[12]. In the early eighteenth century, the Portuguese influence waned and the Sultanates of Oman and later Zanzibar became dominant[4]. In the nineteenth century, the growth of overseas trade, including in enslaved people, led to large-scale population movements from central regions of Africa and settlers from the Yemeni region of Hadramawt[4,13]. In the mid-nineteenth century, Britain and other European powers became

dominant, leading to the settlement of Europeans, the arrival of labourers from South Asia, and further interactions with non-coastal eastern Africans.

In light of this multi-layered history, the extent to which people who identify as Swahili in the present day are genetically linked with people who built the medieval trading towns is unclear, as is the relationship of the medieval groups to earlier groups. Although the intercontinental connections maintained as part of the Indian Ocean trading network meant that foreigners were consistently present along the coast, the extent to which they had families with African residents has long been debated[14].

Swahili traditions suggest that foreigners had an important impact; a set of common oral histories relates the founding of coastal towns to the arrival of a group known as the Shirazi, referring to a region in Persia[3]. This Shirazi tradition was put into writing in the Kilwa Chronicle in the sixteenth century[15]. These accounts of Shirazi roots were central to the narrative constructed by mid-twentieth century colonialist archaeologists, who interpreted second millennium coastal eastern African sites as having been built by Persian and Arab settlers, and focused on connections with the broader Indian Ocean world[16].

However, narratives of foreign origin have the potential to be misleading, as Swahili social ‘elites’ used claims of foreign origin and rejection of cultural connections within Africa to establish their social status and to signal their religious and cultural affinities[17,18].

Recent research has shown that archaeology during colonial times tended to ignore the evidence of deep African roots, emphasizing foreign objects at medieval Swahili sites rather than providing a balanced picture of the archaeological record[2]. Imports at most coastal sites typically comprise less than 5% of total assemblages[2,9]. Other aspects of the material culture also show continuity with earlier settlements, including the persistence of crops, domesticated animals, craft styles and ceramics[9,19]. Linguistic evidence provides additional evidence of African roots: Kiswahili is an African Bantu language with Asian loanwords[20]. However, without ancient DNA evidence, it is not possible to directly address questions of how genetic ancestry changed over time.

We generated ancient DNA data from the skeletal remains of individuals found at six coastal or island towns: Mtwapa, Manda, Faza, Kilwa, Songo Mnara and Lindi. These individuals date to ad 1250–1800 but provide insight into genetic events from the tenth century ad onwards. We also generated ancient DNA from the remains of individuals found at the site of Makwasinyi (postdating around ad 1650), about 100 km inland from the southern Kenyan coast, which was inhabited by people who were in cultural contact with coastal groups. We compare the newly reported data from the ancient individuals with that of present-day coastal Swahili speakers and with previously published data from diverse ancient and present-day eastern African and Eurasian groups.

Dataset overview

We generated 179 ancient DNA libraries from 156 distinct skeletal samples (Supplementary Data File 1 and Methods). We applied in-solution enrichment for a targeted set of about 1.2 million single nucleotide polymorphisms (SNPs) to obtain genome-wide data passing standard measures of ancient DNA authenticity from 80 distinct individuals. The individuals were buried at seven second millennium ad sites in eastern Africa (black shapes in Fig. 1a; individuals are listed in Extended Data Table 1 and Supplementary Data File 2; see Methods, 'Inclusion and ethics' and Supplementary Information for details of archaeological and genetic permissions and sampling). We obtained direct radiocarbon dates (Supplementary Data File 3) for 33 of the skeletons, and estimated date ranges for the other individuals on the basis of archaeological context or genetic evidence of relatedness to individuals for whom we had direct dates (Extended Data Table 1 and Supplementary Data File 2). Because of the reliance on seafood, old carbon entering the food chain (marine reservoir effects) could mean that the radiocarbon dates of some individuals are older than their true dates. Moreover, differences in the dependence on marine food across the archaeological sites could mean that the relative chronology of the coastal individuals and sites may not be possible to determine with full confidence. We also generated new genome-wide data on the Affymetrix Human Origins SNP array from 93 present-day individuals who identified as Swahili and indicated that their ancestors lived for many generations in coastal towns[21] (Supplementary Data File 4). Finally, we generated new

genome-wide data from 19 individuals from Madagascar, and 10 from the United Arab Emirates.

Three sites were northern coastal towns: Mtwapa (Supplementary Fig. 1, 48 individuals spanning ad 1250 to ad 1650), Faza on Pate Island (1 individual), and Manda Island (Supplementary Fig. 2, 8 individuals spanning ad 1450 to ad 1650). Three additional sites were southern coastal towns: Songo Mnara (Supplementary Fig. 3 and Supplementary Table 1, 7 individuals spanning ad 1300 to ad 1800), Lindi (1 individual at ad 1500 to ad 1650), and Kilwa Kisiwani (2 individuals spanning ad 1300 to ad 1600). The remains at Mtwapa, Manda and Songo Mnara were mainly from Muslim burials of elites, often located near mosques. We do not have enough context for the Faza, Kilwa and Lindi burials to know if they followed the same pattern. We also analysed 13 individuals from Makwasinyi (ad 1650–1950), approximately 100 km inland from the coast of present-day Kenya. Although these burials post-date the coastal sites, the Makwasinyi community traded with coastal peoples while remaining isolated in most respects. We hypothesized that the ancestry of Makwasinyi people might be a good proxy to represent inland African groups that may have been in contact with people from medieval towns on the northern Swahili coast in previous centuries[22].

Of the 80 individuals for whom we report data, we exclude 26 in genome-wide analyses, although their data remain valuable. Of these, 18 had too few SNPs for high-resolution whole-genome analyses although they yielded useful data such as reliably determined mitochondrial sequences; 5 were first- or possibly second-degree relatives of other individuals in the dataset with higher-quality data; 2 showed evidence of contamination; and 1 was a population genetic outlier with limited data, raising the possibility of contamination (Supplementary Data File 2).

Ancient DNA data from four individuals from the eastern African coast have previously been published (Supplementary Data File 5), but none have been published from a Swahili town[23]. An individual from around ad 1400 whose remains were recovered from Makangale Cave on Pemba Island had ancestry predominantly related to western African groups[23] (an ancestry common today in speakers of Bantu languages and prevalent in eastern Africa—hereafter referred to as 'Bantu-associated'). Another individual

from Makangale Cave on Pemba Island dated to around ad 600, an individual from around ad 600 from Kuumbi Cave on Zanzibar Island, and an individual from around ad 1500 from Panga ya Saidi in Kenya all had predominantly sub-Saharan African forager-associated ancestry[23]. There is no indication of Eurasian ancestry deriving from migrations in the last 2,000 years in any of these individuals, which differs from nearly all of the individuals from medieval coastal towns newly reported here.

In this Article, we use 'African ancestry' to refer to DNA deriving from people who can be genetically well-proxied by sub-Saharan Africans for whom there is published ancient DNA data dating to between 2000 bc and ad 1000. We use the terms 'Eurasian', 'Persian', 'Arabian' and 'Indian' to refer to ancestry that can be proxied by modern populations from these regions and that are not known to be similar to ancestry in sub-Saharan Africans between 2000 bc and ad 1000. The evidence that a proportion of the ancestors of Africans from between 2000 bc and ad 1000 may have come from Eurasia—for example, approximately 40% of the ancestry of people of the eastern African Pastoral Neolithic culture[24]—does not contradict these definitions, as all humans are mixed at multiple time depths of history. As long as we specify both the time and the geography for the source populations, we can be precise in the use of the term 'African ancestry'[25].

Genetic affinities

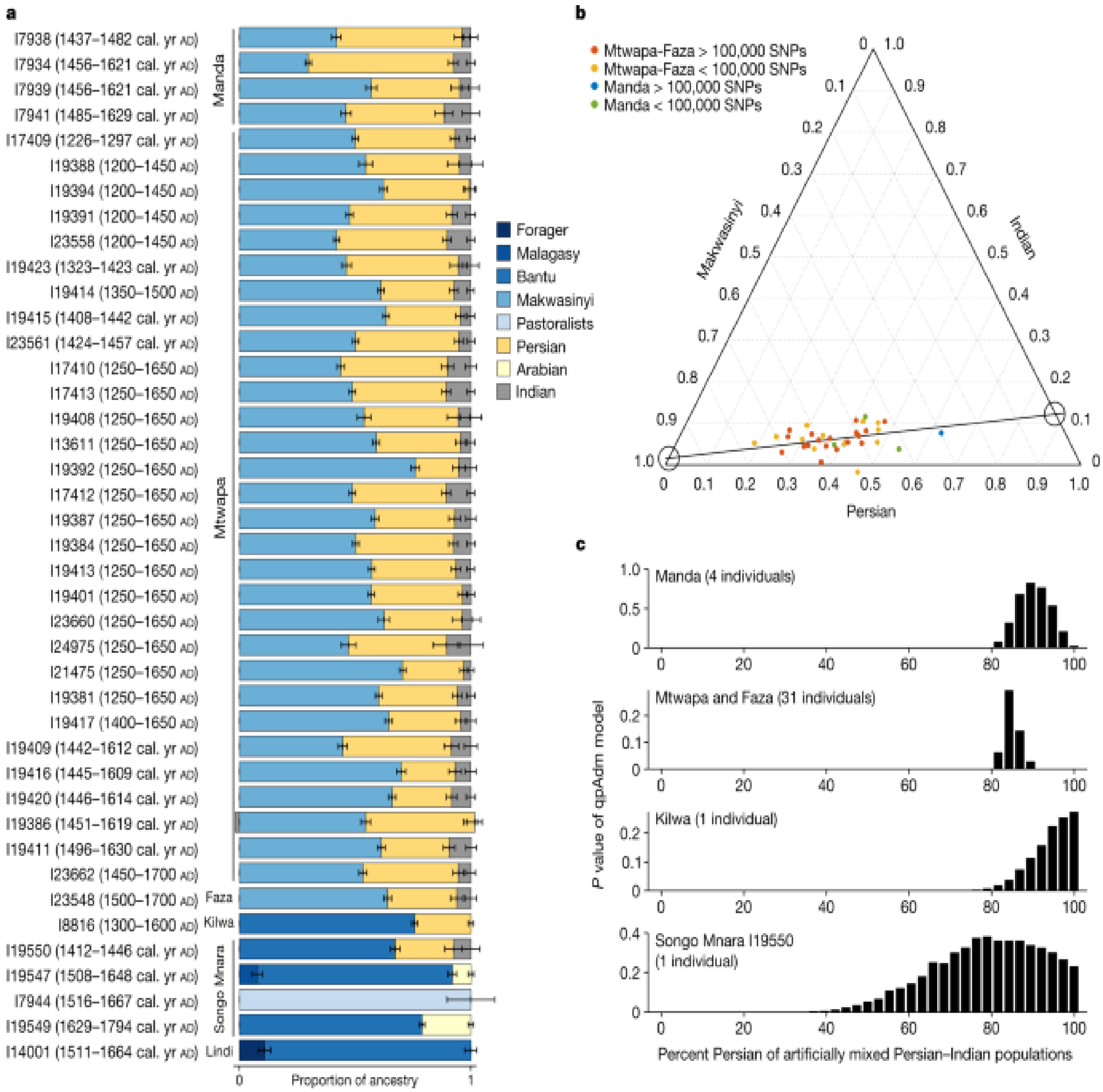
To obtain a qualitative picture of the sources of the ancestry in the ancient individuals, we performed principal component analysis (PCA) (Fig. 1b). We used 1,286 present-day Eurasian and African individuals to compute the axes (Supplementary Data File 6). We projected the newly reported ancient individuals onto this PCA and found that they form a cline, with one end overlapping with ancient and present-day African groups and the other falling between present-day Persians and Indians. This suggests mixtures of different proportions of source populations at either end of the cline, with these sources potentially having multiple deeper ancestry components. Some coastal individuals—particularly from Songo Mnara and Lindi, do not fall on this cline—suggesting additional complexity, although our power to understand this variability is limited by the small sample sizes. Similar patterns are evident with unsupervised clustering using ADMIXTURE,

which further suggests sub-Saharan African-associated components, southwest Asian-associated components, and East Asian or Indian-associated components (Fig. 1c and Extended Data Fig. 1).

Proportions of African, Persian and Indian DNA

Using qpAdm[26], we find that most medieval and early modern individuals can only be fit by a model with at least three ancestry components that can be proxied with ancient African, present-day Iranian and present-day Indian populations (Fig. 2a, Extended Data Table 2, Supplementary Tables 2–10 and Supplementary Information). Such a three-source model fits the pool of 48 Mtwapa individuals and the Faza individual (P value for fit = 0.23); the pool of Manda ($P = 0.28$) individuals, and at least one Songo Mnara individual (I19550) ($P = 0.38$). One Kilwa individual (I8816) had a relatively high proportion of ancestry related to inland sub-Saharan Africans, and so the Indian proportion, which is the smallest contributor, falls below the threshold of definitive detection (a two-source model fits; $P = 0.27$).

Fig. 2: Individual ancestry proportions.

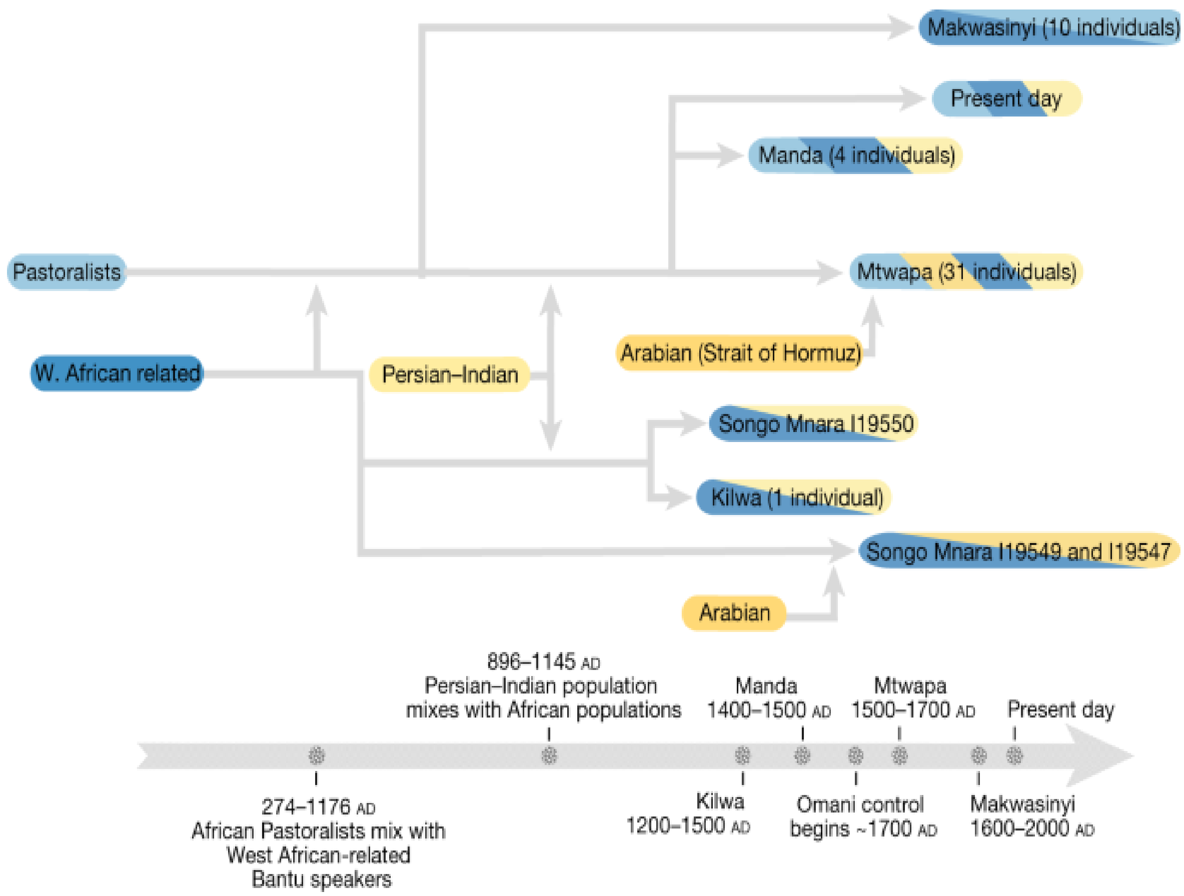


a, Inferences from qpAdm (see Extended Data Table 2 and Supplementary Information for model details and statistical fit). Blue represents African ancestry: the most common are Bantu-associated (common at southern sites) and Makwasinyi associated (northern sites), which itself is approximately 80% Bantu-related and 20% pastoralist-related. Yellow represents Southwest Asian ancestry: Persian or Arabian. Grey represents Indian ancestry. Bars represent s.e.m., computed using a block jackknife across all 5-centimorgan (cM) segments of the autosomes, and are meaningful even for single individuals as the genome contains information from a large sample size of ancestors. **b**, Ternary plot of Makwasinyi, Persian and Indian ancestry components in Mtwapa and Faza (red (high coverage) and yellow (low

coverage)) and Manda (blue (high coverage) and green (low coverage)). Individuals with higher coverage (>100,000 SNPs overlapping positions on the Human Origins SNP array) are used to fit a linear regression (dashed line), which intersects at nearly 100% Makwasinyi and 0% Persian and Indian, consistent with a Makwasinyi-related population with little or no recent Asian ancestry mixing with an already-mixed Persian–Indian population. **c**, Bar graph showing *P* values from Hotelling *T*-squared tests for a qpAdm model with a mixed Persian–Indian source. The *x*-axis specifies the proportion of Persian ancestry in the source.

The type of African ancestry needed to make the models fit differed between individuals from the north (Mtwapa, Faza and Manda) and south (Kilwa and Songo Mnara) of the studied region. In Kenya, the best-fitting proxy African source is the inland Makwasinyi individuals (Extended Data Table 2), who are themselves well-modelled as mixtures of about 80% Bantu-associated and 20% ancient eastern African Pastoral Neolithic ancestry[24] (Fig. 3a, Extended Data Table 3, Supplementary Tables 2 and 3 and Supplementary Information). In Tanzania, the best-fitting African proxy source is Bantu-associated without evidence of a Pastoral Neolithic contribution. We use the individual buried at around ad 1600 in Lindi as a proxy Bantu-associated source for the Kilwa individual and individual I19550 from Songo Mnara (Extended Data Table 2).

Fig. 3: Inferred admixture events along the eastern African coast.



African populations are represented in shades of blue; Southwest Asian and Indian populations are represented in shades of yellow. Populations with both colours represent those that are admixed between the corresponding proxy source populations.

Although three continental sources are required to fit the data, the individuals from Manda, Faza and Mtwapa form a cline in PCA, suggesting two proximal source populations (Fig. 2b). Using linear regression, we extrapolate the ancestry of these two sources and infer that one was consistent with a 100% African origin (Supplementary Fig. 5 and Supplementary Information). The same analysis concludes that the other source had both Persian and Indian ancestry. This is consistent with sub-Saharan Africans mixing with a group that already had a mixture of Persian and Indian ancestry components. Given the two different African sources for the northern and southern individuals, there must have been at least two, but possibly more, admixture events. This would be expected if people of mixed

Persian and Indian ancestry had children with people from different local African populations at different locations along the coast.

When we analyse the Mtwapa and Faza, Manda, Kilwa and Songo Mnara individuals separately, the estimated proportions of Eurasian ancestry from India overlap, which could be consistent with a homogeneous source population of mixed Indian–Persian ancestry for all sites (Fig. 2c). However, a variable proportion among the early immigrants cannot be ruled out, and thus we cannot distinguish between scenarios of two or more streams of Persian–Indian migrants. Our statistical power to detect Indian ancestry relies on pooling of data from multiple individuals. There are a number of individuals with limited data or low Eurasian ancestry, for example individual I8816 from Kilwa, for which we have little power to detect Indian ancestry and cannot definitely document it (Supplementary Information).

Males from Persia and females from Africa

We tested whether male and female ancestors contributed the same proportions of African-like, Persian-like and Indian-like ancestry to ancient individuals in the northern coastal sites and Kilwa (Table 1). To carry out this analysis, we used the fact that chromosomes 1–22, the X chromosome, mitochondrial DNA (mtDNA) and the Y chromosome are passed down to subsequent generations in different ways by males and females. We could not perform the same analysis at Songo Mnara because no individuals with high quality data fit the three-way model.

Table 1 Population mixture and evidence of sex bias in three groups from the Swahili coast

We first analysed mtDNA (Extended Data Table 1 and Table 1). Analysing 62 individuals with confidently determined mtDNA haplogroups (including relatives and individuals with low coverage genome-wide data), we find that 59 carry an L* haplogroup, which in the present day is almost entirely restricted to sub-Saharan Africans[27]. The exceptions are a pair of first- or second-degree relatives from Mtwapa carrying M30d1, which in the present day is largely restricted to South Asia[27], and an individual with haplogroup R0+16189, which today is characteristic of Saudi Arabia and the Horn of

Africa[28]. These results are consistent with female ancestry deriving overwhelmingly from African sources.

Analysing male-transmitted Y chromosome DNA, we find that two out of three non-first-degree related males from Manda carry haplogroup J2, and the third carries G2. Both haplogroups are characteristic of Southwest Asia (plausibly Persia) and are largely absent in sub-Saharan Africans[29]. The Kilwa individual also carries J2. Fourteen out of 19 males from Mtwapa have Y chromosome haplogroups in the J family, and two are of the R1a haplogroup, all considered typically non-African. Only 3 out of 19 Mtwapa males, along with the Faza male, are in the E1 family characteristic of sub-Saharan Africa.

We next compared chromosome X, which occurs as two copies in females and one in males and so reflects mostly female history, to the autosomes (chromosomes 1–22), which equally reflect female and male history (Methods and Supplementary Information). Chromosome X estimates of African ancestry are higher than on the autosomes at all sites, providing an independent line of evidence that African ancestry is primarily from females and Persian ancestry is primarily from males (Table 1). Assuming that the mixture occurred over just a few generations, we obtain quantitative estimates of the proportion of African ancestry from females as 100% at Manda, 69–97% at Mtwapa-Faza, and 69–100% at Kilwa (Methods and Table 1). We estimate Persian ancestry at Mtwapa-Faza and Kilwa to be 100%, and at Manda as 90–100%. If the mixture occurred over more generations, we cannot obtain a point estimate, but can nevertheless infer primarily African female and Persian male sources.

Together, these multiple lines of evidence show that the Southwest Asian ancestors of the Mtwapa and Faza, Manda and Kilwa individuals were almost entirely male, whereas the African ancestors were almost entirely female.

Mixing began by ad 1000

We estimated when mixture occurred on the basis of the sizes of stretches of ancestry inherited from the ancestral populations, which break up at a known

rate every generation[30]. We calculated 95% confidence intervals for the inferred dates of ad 795–1085 for a pool of the northern Mtwapa, Faza and Manda individuals, and ad 708–1219 for a pool of the southern Kilwa and Songo Mnara I19550 individuals (Table 1 and Extended Data Fig. 3). The uncertainty intervals overlap from ad 795 to ad 1085. These estimates would be biased too old if there was a marine reservoir effect. The inferred dates also reflect an assumption that the mixture occurred all at once; however, mixture of Eurasians and Africans was certainly drawn out over a number of generations, and indeed the historical evidence and our genetic analysis that follows document continued incorporation of migrants from both inland Africa and Eurasia until the present. However, simulations show that mixture must have begun by the inferred date[31], and thus we can be confident that already-mixed males with both Indian and Persian ancestry were present along the coast by around ad 1000, and began mixing with primarily female sub-Saharan Africans by that time.

Arabians and other migratory influences

Although almost all the coastal individuals we analysed had Asian ancestry, there were exceptions. Some early modern individuals at Lindi and Songo Mnara showed no evidence of recent Asian ancestry (I14001 and I7944) (Extended Data Fig. 2a, Extended Data Table 3, Supplementary Table 11 and Supplementary Information). We find possible Malagasy-associated ancestry in Songo Mnara (I19547) (Extended Data Fig. 2a, Extended Data Table 3 and Supplementary Information). Our finding of coastal individuals who differ from others from similar times or regions attest to continued exchange with people in the Indian Ocean trading network, although our sample size is too small to identify general patterns.

For some of the individuals in our study with Asian ancestry (from Manda, Kilwa and I19550 from Songo Mnara), there is only evidence of Persian or Persian–Indian ancestry. However, other individuals, particularly from Mtwapa, can only be modelled by sources with some Arabian-associated ancestry when using other Mtwapa or Manda individuals as a source (Supplementary Fig. 6, Supplementary Tables 12–15 and Supplementary Information). We were unable to determine the exact source of the Arabian-associated ancestry. However, we know that it is somewhere on the genetic

gradient between Arabians and Persians. A proxy source that provides a fit for this Arabian-related ancestry at Mtwapa is present-day people who live on the shores of the Strait of Hormuz, which separates the Arabian Peninsula from Iran. The Strait of Hormuz and the Swahili coast were under Omani control by the end of the seventeenth century.

Direct genetic evidence for Arabian-associated migration comes from two individuals from Songo Mnara. Both date to the early modern period when contacts with Arabia are well documented, and they can only be modelled with Arabian-related ancestry in qpAdm (Fig. 2a). Analyses of present-day coastal populations also point to Arabian genetic influences. Whereas the Asian ancestry of some individuals can be modelled entirely as Persian–Indian as in medieval Manda, for other individuals, ancestry from the Strait of Hormuz is a better fit (coloured salmon in Supplementary Data File 7).

Relating medieval to modern Swahili

We assembled genome-wide data for two groups of present-day people who identify as Swahili: 89 with previously reported data[32] and 93 for which we generated new data.

Most of the individuals in the previously published dataset (87 out of 89) have only a modest inheritance from people with ancestry resembling the medieval people that we sampled from nearby coastal areas (Extended Data Fig. 2a, Extended Data Table 3, Supplementary Table 16 and Supplementary Information). We estimate that 11 ± 4% have medieval Swahili-associated ancestry; 84 ± 3% have Bantu-associated ancestry; and 6 ± 3% have Pastoral Iron Age-associated ancestry[24] (Extended Data Fig. 2a and Extended Data Table 3). These patterns are mirrored in the Y chromosome haplogroups, with 95% having typically African haplogroups—in contrast to the medieval coastal individuals, for whom almost all Y haplogroups are associated with Near Eastern people (Tables 1 and 2). However, the newly generated data show a much higher inferred proportion of ancestry from groups similar to the medieval ones, ranging from 46 to 77% medieval Swahili-associated ancestry (depending on whether we use a pool of Manda and Mtwapa individuals as the sources) (Extended Data Table 4 and Supplementary Data File 7). Y chromosome haplogroups in the new data are also consistent with

a greater contribution from medieval peoples: the African-associated Y-haplogroup frequency was 45%, much larger than the 17% in the medieval individuals, but smaller than the 95% estimated in the previously published study (Table 2).

Table 2 Comparison of haplogroup distributions in medieval Swahili individuals with those from two studies of present-day Swahili people

The differences in the proportions of ancestry in the present-day individuals identifying as Swahili may reflect differences in how they were estimated. The previously published dataset included people from the coastal towns of Kilifi, Lamu and Mombasa in Kenya who indicated that their family had been Swahili-speaking for the past three generations. The newly published dataset included people from 13 locations along the Kenyan coast who indicated that their ancestors had lived in coastal towns and had a Swahili identity for many generations, thus enriching for more traditional upper-class Swahili people who plausibly retained more ancestry from medieval coastal individuals[30]. The greater medieval coastal ancestry may also reflect isolation. In the newly published dataset from the mainland, individuals from the site of Jomvu Kuu had significantly less medieval coastal ancestry than the individuals from the other sites, all of which were from islands that were plausibly more isolated from admixture with inland groups ($P = 3.3 \times 10^{-6}$, one-sided Wilcoxon rank sum test).

We used the lengths of African and Persian genetic segments to estimate the age of admixture in the newly published modern individuals[30] at ad 1096–1410. These data are more recent than the medieval coastal samples, consistent with ongoing mixture with African or Asian populations since medieval times.

No recent Asian ancestry in inland people

The Makwasinyi individuals date to the past three centuries, and are from deep in the Tsavo region, nearly 50 km from the nearest population centre. The Makwasinyi group fits as a proxy source for African ancestry in qpAdm modelling of the Mtwapa, Faza and Manda individuals, but unlike these individuals, qpAdm finds no evidence of recent Asian ancestry in the

Makwasinyi group (similar to present-day non-coastal populations; Supplementary Table 17). Instead, Makwasinyi individuals are similar in ancestry to the modern individuals identified as Swahili in the previously reported dataset[32]. They are well-modelled as 21.3 ± 1.2% Pastoral Neolithic-associated ancestry (from herders present in eastern African after 3000 bc) and 78.7 ± 1.2% Bantu-associated ancestry (from farmers present after 1000 bc) (Extended Data Fig. 2a). We did not detect sex bias in the history of the formation of this population (Methods, Supplementary Information). The average date for the Bantu-Pastoral Neolithic-associated mixture is around ad 300–1200 (Extended Data Fig. 3), with most of this range consistent with the archaeological evidence for the impact of the Bantu expansion on this region.

Discussion

A key finding of this study is evidence of mixture at roughly ad 1000 between peoples of African and Persian ancestries (Fig. 3 and Table 1). This is consistent with the Kilwa Chronicle, which describes the arrival of Persians on the Swahili coast and interactions between them and coastal residents. Whether or not this history has a basis in an actual voyage, the ancient DNA provides direct evidence for Persian-associated ancestry being derived overwhelmingly from males and arriving on the eastern African coast by about ad 1000. This timing coincides with archaeological evidence for a substantial cultural transformation on the coast, including the widespread adoption of Islam[10]. At Kilwa, coin evidence has dated a ruler linked to a Shirazi (Persian) dynasty, Ali bin al-Hasan, to the mid-eleventh century[33]. The genetic evidence suggests that this arrival was accompanied by mixture, which began by ad 1000, and continued later. People of both African and Asian ancestry made major contributions, with African proportions of approximately 57% on average at Mtwapa and Faza, 32% at Manda, 67% at Songo Mnara and 74% at Kilwa (Table 1, Extended Data Table 2 and Supplementary Table 8).

Archaeological evidence provides important context for our genetic findings. The individuals that we analysed lived in the thirteenth to eighteenth centuries, and were excavated mostly from elite contexts. However, coastal sites from around ad 1000, when the mixture occurred,

showed little evidence for distinct societal elites. Three of the sites sampled here (Mtwapa, Songo Mnara and Faza) did not exist as towns in ad 1000, and so these admixed populations moved to those towns later. Thus, the elite inhabitants of Mtwapa and other sites developed from admixed populations and were not foreign migrants or colonists.

Linguistic evidence provides further context. Kiswahili is a Bantu language, and since most ancestry in medieval Swahili people derives from African people, our results suggest that the children of immigrant men of Asian origin adopted the languages of their mothers, a common pattern in matrilocal cultures[34]. However, Kiswahili also has non-African influences, reflecting a millennium and a half of intense interaction with societies around the Indian Ocean rim. Persian loanwords contribute up to 3% of Kiswahili, but it is unclear whether they are derived directly from Persian or through adoption into other Indian Ocean languages[35]. Arabic loanwords are the single largest non-Bantu element in Kiswahili[35] (16–20% of words), and may be primarily due[20] to incorporations after ad 1500.

A recurrent theme of our findings is the different participation of males and females in population mixture events. We find evidence of predominantly male Southwest Asian ancestors mixing with predominantly female African and, to a much lesser extent, female Indian ancestors in the lineages of medieval people on the Swahili coast. This provides evidence for asymmetric social interactions between groups as cultural contact occurred, although such genetic data cannot reveal the processes contributing to these patterns.

This study provides information from only a subset of times and places relevant to the medieval coastal civilization, and it is important to recognize these biases. The geographical coverage is skewed towards Kenya, with the individuals from Tanzanian sites such as Songo Mnara, Kilwa and Lindi being sufficient to identify similarities and differences in ancestry profiles from Kenya but not sufficient to define a general pattern. In addition, the individuals that we analysed were not fully representative of all social and economic groups in Swahili society. Nearly all of the coastal graves and tombs in this study occupied prominent positions in medieval and early modern townscapes (see Supplementary Information). With the possible

exceptions of Kilwa, Lindi and Faza, we analysed elite individuals from high-profile coastal sites. However, the Swahili cultural world included many non-elite settlements, where ancestry might be systematically different[36]. Our analysis of data from two samplings of present-day people who identify as Swahili with different strategies for determining this identity also reveals qualitative differences in ancestry patterns, revealing how groups identified as Swahili retain high substructure and variation today.

These findings highlight multiple directions for future work on ancient DNA. One approach is to study individuals pre-dating the twelfth century, including before and after the major population mixtures that we show occurred around ad 1000. Another approach is to study individuals from unsampled parts of the Swahili world, including the present-day countries of Somalia, Mozambique, the Comoros Islands and Madagascar. However, the results presented here provide unambiguous evidence of ongoing cultural mixing on the East African coast for more than a millennium, in which African people interacted and had families with immigrants from other parts of Africa and the Indian Ocean world. Narratives of ancestry on the eastern African coast have a complex history, and the genetic findings of long-standing, sex-biased mixtures add to this complexity.

Methods

Inclusion and ethics

The present-day communities of the Swahili coast have strong traditions of connection to the people of the medieval coastal towns—including in some cases a tradition of descent from the people who lived in these communities, as well as shared language and religion—and thus community consultation is an important part of this work. Many medieval and present-day coastal people are also Muslims, and thus it is important to carry out any analysis in ways that are sensitive to Muslim proscriptions against disturbing the dead. The sampling for this study emerged from decades-long community-based archaeology projects—including by S.K. and C.M.K. in the Kenya region, and by S.W.-J., J.F., Tanzanian Antiquities and the Songo Mnara Ruins Committee in the Tanzania region—that involved participation in and return of results to local communities, following rules for handling and reburying

of remains agreed by the communities. Prior to submission of this study for publication, the corresponding authors held return-of-results consultation meetings in Lamu, Songo Mnara and Kilwa Kisiwani, and feedback from these engagements was incorporated into the final manuscript. The community meeting held in Lamu was both in person and on zoom. Mtwapa no longer has a viable Swahili community. The descendants of the residents of Manda are no longer visible, but given Manda's proximity to Lamu, we chose Lamu as the most viable community to host the forum. Makwasinyi elders were consulted in 2001 and 2002 during fieldwork and again to discuss the results of the present analysis.

Ancient DNA data generation

To generate the genetic data for the samples in Extended Data Table 1 and Supplementary Data File 2 (by-library metrics in Supplementary Data File 1), we used established protocols in dedicated clean rooms, involving first sampling typically 40 mg of powder from skeletal remains, then extracting DNA using methods designed to retain short, damaged fragments[37], then building individually barcoded libraries[38] after incubation with uracil DNA glycosylase (UDG) with the goal of greatly reducing the characteristic errors typical of ancient DNA. We amplified libraries and carried out in-solution enrichment of them for about 1.2 million SNPs[39] as well as baits targeting the mitochondrial genome[40]. We sequenced the enriched products along with a small amount of unenriched library on Illumina NextSeq 500 or HiSeqX10 instruments. The resulting sequenced paired-end reads were separated into respective libraries and stripped of identification tags and adapter related sequences. Read pairs were merged prior to alignment, requiring a minimum of 15 base pair overlap, allowing one mismatch if base quality was at least 20, or up to three mismatches at lower base quality. We restricted to sequences that were at least 30 base pairs long. The resulting sequences were aligned using samse from bwa-v0.6.1[41] separately to both the human genome reference sequence (hg19) and the mitochondrial RSRS genome[42] to allow the targeted nuclear and mitochondrial specific sequences to be assessed.

We performed a series of quality control measures. We inferred the rate of mismatch of the sequences mapping to the mitochondrial consensus

sequence (of libraries with at least twofold coverage) and flagged libraries where the upper bound of the inferred match rate[43] was at least 95%; we also examined the rate of polymorphism on the X chromosome in males and flagged individuals as appreciably contaminated if the lower bound of the 95% confidence interval[44] was at least 1%. We tested if the ratio of Y chromosome to the sum of X and Y chromosome sequences was in the range expected for DNA entirely from a female (<3%) or male (>35%), and we tested for an elevated rate of cytosine to thymine substitution in the final nucleotide using a threshold of at least 1.8% (this was a lower threshold than is typical in ancient DNA studies due to the young age of the samples; the median damage rate of libraries is 9.5%). If a sample had sufficient evidence of contamination, we restricted to only damaged molecules of sequenced DNA, and these samples are denoted with '_d' in their Genetic ID in Supplementary Data File 2. Several samples did not pass quality control; these are any samples where all libraries failed in Supplementary Data File 1. No individual from two sampled sites—Bungule and Shanga—passed quality control. For each individual, we represented each position in the genome based on a single randomly selected sequence

A total of 80 individuals passed quality control screening. However, the analyses in this study focus on the 54 non-outlier individuals that are not first- or second-degree relatives of another individual in the study for which we have better quality data and for which we have data on a minimum of 15,000 SNPs. Relative relationships are determined by a method similar to READ[45].

Supplementary Data File 1 provides a full report of sequencing results on both libraries that yielded data passing standard measures of ancient DNA authenticity and libraries that failed screening; we also report negative results as a guide for future sampling efforts.

Genotyping of modern Swahili individuals

We genotyped 93 present-day Swahili individuals, originally collected in ref. [21], on the Affymetrix Human Origins SNP array[26]. As described in the study where mtDNA and Y chromosome haplogroup information for these individuals was initially reported[21], individuals known to the research team

to have origins in or near the targeted recruitment communities assisted in participant recruitment in a form of ‘snowball sampling’. On sampling days, potential participants belonging to families known to have long residency in the community were approached and, if they consented to participation, would be asked afterwards if they knew of anyone else not closely related to them who might be available that day who met the inclusion criteria of being a healthy individual of at least 18 years of age who is a resident of that town and whose grandparents all identify as belonging to a group typically included in the broader Swahili identity. Each participant was provided information about the project in either English or Swahili from Institutional Review Board (IRB)-approved forms and those who elected to continue provided written consent. Protocols for collecting saliva and genealogical information were reviewed and approved by the Lehman College (no. 141-09-070) and City University of New York Integrated IRB (no. 323935). The saliva samples and genealogical information were collected in December 2009 and January 2010 in the Kenyan towns of Faza, Jomvu Kuu, Kizingitini, Lamu, Mikindani, Ndau, Pate, Siyu, Tchundwa and Wasini. A total of 96 samples were chosen for whole-genome analysis to maximize geographic spread, with 16 individuals each from Faza, Kizingitini, Ndau, Pate, Wasini and Jomvu Kuu. We limited to individuals who self-reported that the origin of both their parents and grandparents are local and Swahili, with a priority for males. Data from a total of 93 individuals passed quality control and their data are reported and analysed.

We genotyped on the Human Origins SNP array 19 newly reported individuals from Madagascar, for whom samples were collected between 2007–2014 with approval by the Human Subjects’ Ethics Committees of the Health Ministry of Madagascar and by committees in France (Ministry of Research, National Commission for Data Protection and Liberties and Persons Protection Committee). Individuals all gave written consent before the study. Samples came from two locations in the north and south of the country. Sampled villages were founded before 1900, and individuals were 61 ± 15 years old, with the maternal grandmother and paternal grandfather born within 50 km of the sampling location.

We genotyped on the Human Origins SNP array 10 newly reported Emirati samples that were collected from Emirati nationals from the city of Al Ain, United Arab Emirates. Ethical approval was obtained from the Al Ain

District Human Research Ethical Committee. The samples were collected from healthy adults, and first genotyped on another SNP array as controls for a rare disease study[46].

A list of the newly reported individuals with data on the Human Origins SNP array and some relevant information can be found in Supplementary Data File 4.

Principal component analysis

We used smartpca version 18180 from EIGENSOFT version 8.0.0[47] with optional parameters (numoutlieriter: 2, numoutevec: 3, lsqproject: YES, newshrink: YES, and hiprecision: YES). We computed eigenvectors of the covariance matrix of SNPs from present-day individuals from Africa and Eurasia genotyped on the Affymetrix Human Origins (HO) SNP array, which targets approximately 600,000 SNPs that are a subset of those targeted by in-solution enrichment (see Supplementary Information for a list of populations used). Ancient individuals and present-day individuals (most genotyped on the HO SNP array, and some also genotyped on the Illumina Human Omni5 Bead Chip[32]) were projected into the three-dimensional (3D) eigenspace determined by the first three principal components.

ADMIXTURE clustering analysis

We prepared data for the ADMIXTURE[48] plots in Fig. 1c and Extended Data Fig. 1 using PLINK2[49]. We applied the maf 0.01 option, which only includes SNPs with minor allele frequency of at least 0.01. We pruned SNPs based on LD, using the indep-pairwise option with a window size of 200 variants, a step size of 25 variants, and a pairwise r^2 threshold of 0.4. We ran 4 replicates of ADMIXTURE with random seeds with $K = 4$ to $K = 12$ ancestral reference populations.

Estimates of mixture proportions

We used qpAdm (ADMIXTOOLS version 7.0.2)[26] to test if the Makwasinyi, Manda, Mtwapa and Faza, Kilwa, and Songo Mnara target populations can be formally modelled as having derived from a set of source populations

(termed ‘left’ groups) relative to a set of reference populations (termed ‘right’ groups). See Supplementary Information for further details. qpAdm also provides a *P* value for fit to the model based on a block jackknife across all 5 cM segments of the autosomes.

In our application of qpAdm, we use a cycling approach, treating the target as a linear combination of all possible subsets of the candidate source populations, and moving the other candidate source population to the right. Cycling populations to the right allows us to test if a proposed set of left source populations is consistent with being more closely related to the target than other populations. Thus, we can build the closest admixture model within the constraints of our dataset and test its fit to the data.

Y and mitochondrial haplogroups

Mitochondrial haplogroups were determined using HaploGrep 2[50]. Y chromosome haplogroups were determined according to the Yfull tree version 8.09 (https://github.com/YFullTeam/YTree/blob/master/ytree/tree_8.09.0.json)[51].

Sex bias from X-autosome comparisons

We estimated the extent to which the ancestry from each source population was contributed by female and male ancestors. To do this, we compared the inferences of proportions of ancestry on chromosomes 1–22 (the autosomes) which reflect 50% female and 50% male inheritance (the autosomal coefficient for proportion of a specific ancestry, $C_A = (m + f)/2$), and on chromosome X, which reflects 67% female and 33% male ancestry (the X chromosome coefficient for proportion of a specific ancestry, $C_X = (2f + m)/3$). We determined *Z*-scores for differences between these two estimates as in ref. [52]. For each of Mtwapa (including the one Faza individual), Manda, Kilwa and Makwasinyi, we quantified sex bias in the following manner. (1) We sampled 10^6 sets of coefficients by generating random numbers from a multivariate normal distribution based on the qpAdm-determined jackknife mean coefficients and the error covariance matrix for both the autosomes and chromosome X, separately. We ensured that the eigenvalues of the matrices are all greater than or equal to zero by adding a

small offset to the matrices determined as the absolute value of its respective minimum eigenvalue. (2) We removed from consideration all infeasible sets of coefficients, namely those that include a coefficient below 0. (3) We next calculated the female and male proportions of ancestry for each source population. Given the coefficient C_A on the autosomes and the coefficient C_X on the X chromosome, we can determine the likely proportion of female and male ancestry from any given source by solving the system of two above equations, giving $f = 3C_X - 2C_A$ and $m = 4C_A - 3C_X$. We calculate these f and m proportions for each source population for the sampled set of autosomal and X chromosome coefficients. (4) We normalize the proportions of female and male ancestry as $\hat{f}=f/(f+m)$ and $\hat{m}=m/\left(f+m\right)$ for each source for the sampled sets. (5) We found the mean and standard deviation of all the $\hat{f}$ and $\hat{m}$ values from the sampled set for each of the source populations. (3) We took 95% confidence intervals of all the sampled sets of ancestry proportions and record the values as ranges in Table 1. Under the simple model of all the mixture occurring at once, these formulae can be interpreted as estimating the fraction of ancestors at the time of mixing that are on the male or female side. However, if the mixture was more gradual, the interpretation is more complicated albeit still informative about sex bias.

Proximate sources of the coastal cline

We plot all the Mtwapa, Faza and Manda individuals on a ternary plot with their respective proportions of Makwasinyi, Iranian and Indian ancestry. We apply a linear regression to the high coverage individuals from Mtwapa (I19381, I19394, I19384, I19414, I19420, I19413, I19417, I19401, I23662, I17409, I19391, I13611, I17412, I17413, I23558, I19415, I23561, and I21475) and Manda (I7934) as seen in the ternary plot (see Fig. 2b and Supplementary Fig. 3). The associated Cartesian equation is $f(x) = 0.09856x + 0.01301$. The line intersects the Makwasinyi axis at 98 ± 6% (left circle in Fig. 2b and Supplementary Fig. 3), allowing for 2 ± 6% Indian ancestry. This is statistically consistent with 100% Makwasinyi and 0% Iranian and 0% Indian ancestries. The other end of the regression line intersects the Indian axis (right circle in Fig. 2b and Supplementary Fig. 3) at 12 ± 5% Indian ancestry and 88 ± 5% Iranian ancestry. We further refined these proportions by jackknifing the coefficient estimates. We estimate that one admixing

source population is 98.4 ± 2.5% Makwasinyi-related, consistent with a parsimonious model of 100% Makwasinyi ancestry at that end of the cline. The other source population is 12.2 ± 2.8% Indian and 87.8 ± 2.8% Iranian (see Supplementary Information).

Determining the date of mixture

DATES measures the covariance between pairs of positions in the genome separated by a specified genetic distance, for an admixture model with two source populations[30,31]. This analysis is applied separately to all individuals, and the inferences are pooled across individuals to increase resolution. We pool the individuals from Mtwapa, Faza, and Manda into a northern group, and the individuals from Kilwa and Songo Mnara I19550 into a southern group. We use a 28 ± 2 year-to-generation conversion estimate[53] to calculate the average date of the admixture.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Bam files of aligned reads for the 80 newly published ancient individuals can be obtained from the European Nucleotide Archive (accession no. PRJEB58698). SNP array genotype data for 122 newly reported modern individuals can be obtained from a permanent link at the Dataverse repository at (https://doi.org/10.7910/DVN/NC28XW). The data for 103 of these individuals can be downloaded without registering, whereas the informed consent for 19 individuals from Madagascar is not consistent with unmediated public posting of data, and may be downloaded after filling out a form with an email address and institutional or professional affiliation, and including an affirmation of the following statements: (a) I will not distribute the data outside my collaboration; (b) I will not post the data publicly; (c) I will make no attempt to connect the genetic data to personal identifiers for

the samples; (d) I will use the data only for studies of population history; and (e) I will not use the data for commercial purposes.

References

1. Wynne-Jones, S. & LaViolette, A. J. *The Swahili World* (Routledge, 2017).

2. Kusimba, C. M. *The Rise and Fall of Swahili States* (Altamira Press, 1999).

3. Spear, T. T. The Shirazi in Swahili traditions, culture and history. *Hist. Afr.* **11**, 291–305 (1984).

4. Sheriff, A. *Slaves, Spices and Ivory in Zanzibar. Integration of an East African Commercial Empire into the World Economy, 1770–1873* (James Currey, 1987).

5. Spear, T. Early Swahili history reconsidered. *Int. J. Afr. Hist. Stud.* **33**, 257–290 (2000).

6. Kusimba, C. M. in *State Formations: Global Histories and Cultures of Statehood* (eds Anderson, G., Brooke, J. L. & Strauss, J. C.) 90–107 (Cambridge Univ. Press, 2018).

7. Fleisher, J. & Wynne-Jones, S. Ceramics and the early Swahili: deconstructing the early Tana tradition. *Afr. Archaeol. Rev.* **28**, 245–278 (2011).

8. Pollard, E. & Kinyera, O. C. The Swahili Coast and the Indian Ocean trade patterns in the 7th–10th centuries ce. *J. S. Afr. Stud.* **43**, 927–947 (2017).

9. Horton, M. C. *Shanga: An Early Muslim Trading Community on the Coast of Eastern Africa*, Vol. 14. (British Institute in Eastern Africa, 1996).

10. Fleisher, J. et al. When did the Swahili become maritime? *Am. Anthropol.* **117**, 100–115 (2015).

11. Wynne-Jones, S. *A Material Culture: Consumption and Materiality on the Coast of Precolonial East Africa* (Oxford Univ. Press, 2016).

12. Vernet, T. Le territoire hors les murs des cités-états Swahili de l'Archipel de Lamu, 1600–1800. *J. Afr.* **74**, 381–411 (2004).

13. Glassman, J. *Feasts and Riot: Revelry, Rebellion, and Popular Consciousness on the Swahili Coast* 1856–1888 (James Currey, 1995).

14. Caplan, P. 'But the coast, of course, is quite different': academic and local Ideas about the East African littoral. *J. East. Afr. Stud.* **1**, 305–320 (2007).

15. Freeman-Grenville, G. S. P. *The East African Coast. Select Documents from the First to the Earlier Nineteenth Centuries* (Clarendon Press, 1962).

16. Chittick, H. N. in *East Africa and the Orient: Cultural Syntheses in Pre-Colonial Times* eds Chittick, H. N. & Rotberg, R. I.) 16–43 (Africana, 1975).

17. Pouwels, R. L. *Horn and Crescent. Cultural change and traditional Islam on the East African coast,* 800–1900. *African Studies Series*, Vol. 53. (Cambridge Univ. Press, 1987).

18. Saad, E. Kilwa dynastic historiography: a critical study. *Hist. Afr.* **6**, 177–209 (1979).

19. Crowther, A. et al. Subsistence mosaics, forager-farmer interactions, and the transition to food production in eastern Africa. *Quat. Int.* **489**, 101–120 (2018).

20. Nurse, D. & Spear, T. *The Swahili: Reconstructing the History and Language of an African Society,* 800–1500 (Univ. of Pennsylvania Press, 1985).

21. Raaum, R. L. et al. in *The Swahili World* (eds Wynne-Jones, S. & LaViolette, A. J.) 81–102 (Routledge, 2017).

22. Kusimba, C. M., Kusimba, S. B. & Dussubieux, L. Beyond the coastalscapes: preindustrial social and political networks in East Africa. *Afr. Archaeol. Rev.* **30**, 399–426 (2013).

23. Skoglund, P. et al. Reconstructing prehistoric African population structure. *Cell* **171**, 59–71.e21 (2017).

24. Prendergast Mary, E. et al. Ancient DNA reveals a multistep spread of the first herders into sub-Saharan Africa. *Science* **365**, eaaw6275 (2019).

25. Mathieson, I. & Scally, A. What is ancestry? *PLoS Genet.* **16**, e1008624 (2020).

26. Patterson, N. et al. Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).

27. Quintana-Murci, L. et al. Where West meets East: the complex mtDNA landscape of the Southwest and Central Asian Corridor. *Am. J. Hum. Genet.* **74**, 827–845 (2004).

28. Gandini, F. et al. Mapping human dispersals into the Horn of Africa from Arabian Ice Age refugia using mitogenomes. *Sci. Rep.* **6**, 25472 (2016).

29. Cadenas, A. M. et al. Y-chromosome diversity characterizes the Gulf of Oman. *Eur. J. Hum. Genet.* **16**, 374–386 (2008).

30. Chintalapati, M., Patterson, N. & Moorjani, P. The spatiotemporal patterns of major human admixture events during the European Holocene. *eLife* **11**, e77625 (2022).

31. Narasimhan Vagheesh, M. et al. The formation of human populations in South and Central Asia. *Science* **365**, eaat7487 (2019).

32. Brucato, N. et al. The Comoros show the earliest Austronesian gene flow into the Swahili Corridor. *Am. J. Hum. Genet.* **102**, 58–68 (2018).

33. Horton, M., Olsen, J., Fleisher, J. & Wynne-Jones, S. The chronology of Kilwa Kisiwani, ad 800–1500. *Afr. Archaeol. Rev.* **39**, 143–166 (2022).

34. Lansing, J. S. et al. Kinship structures create persistent channels for language transmission. *Proc. Natl Acad. Sci. USA* **114**, 12910–12915 (2017).

35. Haspelmath, M. & Tadmor, U. *Loanwords in the World's Languages : A Comparative Handbook* (De Gruyter Mouton, 2009).

36. LaViolette, A. & Fleisher, J. B. Developments in rural life on the Eastern African coast, a.d. 700–1500. *J. Field Archaeol.* **43**, 380–398 (2018).

37. Dabney, J. et al. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl Acad. Sci. USA* **110**, 15758–15763 (2013).

38. Rohland, N. et al. Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. *Philos. Trans. R. Soc. Lond. B* **370**, 20130624–20130624 (2015).

39. Fu, Q. et al. An early modern human from Romania with a recent Neanderthal ancestor. *Nature* **524**, 216–219 (2015).

40. Maricic, T., Whitten, M. & Pääbo, S. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS ONE* **5**, e14004–e14004 (2010).

41. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).

42. Behar, DoronM. et al. A "Copernican" reassessment of the human mitochondrial DNA tree from its root. *Am. J. Hum. Genet.* **90**, 675–684 (2012).

43. Fu, Q. et al. DNA analysis of an early modern human from Tianyuan Cave, China. *Proc. Natl Acad. Sci. USA* **110**, 2223–2227 (2013).

44. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics* **15**, 356 (2014).

45. Monroy Kuhn, J. M., Jakobsson, M. & Günther, T. Estimating genetic kin relationships in prehistoric populations. *PLoS ONE* **13**, e0195491 (2018).

46. Reiff, R. E. et al. METTL23, a transcriptional partner of GABPA, is essential for human cognition. *Hum. Mol. Genet.* **23**, 3456–3466 (2014).

47. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).

48. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).

49. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**, 7 (2015).

50. Weissensteiner, H. et al. HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res.* **44**, W58–W63 (2016).

51. Lazaridis, I. et al. The genetic history of the Southern Arc: A bridge between West Asia and Europe. *Science* **377**, eabm4247 (2022).

52. Mathieson, I. et al. The genomic history of southeastern Europe. *Nature* **555**, 197–203 (2018).

53. Moorjani, P. et al. A genetic method for dating ancient genomes provides a direct estimate of human generation interval in the last 45,000 years. *Proc. Natl Acad. Sci. USA* **113**, 5652 (2016).

54. Lazaridis, I. et al. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409–413 (2014).

55. Lazaridis, I. et al. Genomic insights into the origin of farming in the ancient Near East. *Nature* **536**, 419–424 (2016).

56. Nakatsuka, N. et al. The promise of discovering population-specific disease-associated genes in South Asia. *Nat. Genet.* **49**, 1403–1407 (2017).

Acknowledgements

We thank N. Adamski, R. Bernardos, B. Culleton, I. Greenslade, D. Kennett, M. Mah, A. Micco and Z. Zhang for contributions to data generation, processing and curation. E.S.B. was supported by EMBO Postdoctoral fellowship ALTF 242-2021. The work of C.M.K., S.B.K., S.R.W. and J.M. in Kenya was supported by the National Museums of Kenya and the Republic of Kenya. The bulk of the research at Manda, Mtwapa, Faza and Makwasinyi was carried out when C.M.K. was at the Field Museum of Natural History. We acknowledge financial support from the US National Science Foundation SBR 9024683 (1991–1993); BCS 9615291 (1996–1998); BCS 0106664 (2002–2004); BCS 0352681 (2003–2004); BCS 0648762 (2007–2009; BCS 1030081 (2010–2012), the US National Endowment for the Humanities (2012–2014), the US IIE J. W. Fulbright Sr Scholars Program 2002–2003; 2012), and the National Geographic Society (1996–1997). The research of S.W.-J. and J.F. at Songo Mnara was supported by the Antiquities Division of the Ministry of Natural Resources and Tourism, Tanzania, and also by the National Science Foundation (BCS 1123091) and the Arts and Humanities Research Council (AH/J502716/1). The Madagascar modern sample collection was supported by the MAGE consortium. D.R. is an Investigator of the Howard Hughes Medical Institute, and the ancient DNA laboratory work and analysis were also supported by National Institutes of Health grant HG012287, by John Templeton Foundation grant 61220, by the Allen Discovery Center program, which is a Paul G. Allen Frontiers Group advised program of the Paul G. Allen Family Foundation, and by a gift from J.-F. Clin. S.W.-J. and all co-authors are grateful to UK Research and Innovation for enabling fully Open Access

publication of this study via financial support to the York Open Access Fund.

Author information

Author notes

1. These authors contributed equally: Jeffrey Fleisher and Stephanie Wynne-Jones

2. These authors jointly supervised this work: David Reich and Chapurukha M. Kusimba

Authors and Affiliations

1. Department of Human Evolutionary Biology, Harvard University, Cambridge, MA, USA

   Esther S. Brielle, Kendra Sirak, Nasreen Broomandkhoshbacht, Kim Callan, Elizabeth Curtis, Lora Iliev, Ann Marie Lawson, Jonas Oppenheimer, Kristin Stewardson, J. Noah Workman, Fatma Zalzala, Nick Patterson & David Reich

2. Department of Anthropology, Rice University, Houston, TX, USA

   Jeffrey Fleisher & Mary E. Prendergast

3. Department of Archaeology, University of York, York, UK

   Stephanie Wynne-Jones

4. University of South Africa, Pretoria, South Africa

   Stephanie Wynne-Jones

5. Department of Genetics, Harvard Medical School, Boston, MA, USA

Kendra Sirak, Nasreen Broomandkhoshbacht, Kim Callan, Elizabeth Curtis, Lora Iliev, Ann Marie Lawson, Jonas Oppenheimer, Lijun Qiu, Kristin Stewardson, J. Noah Workman, Fatma Zalzala, Swapan Mallick, Nadin Rohland, Mary E. Prendergast & David Reich

6. Howard Hughes Medical Institute, Harvard Medical School, Boston, MA, USA

Nasreen Broomandkhoshbacht, Kim Callan, Elizabeth Curtis, Lora Iliev, Ann Marie Lawson, Lijun Qiu, Kristin Stewardson, Fatma Zalzala, Swapan Mallick & David Reich

7. Jaramogi Oginga Odinga University of Science and Technology, Bondo, Kenya

George Ayodo

8. National Museums of Tanzania, Dar es Salaam, Tanzania

Agness O. Gidna & Amandus Kwekason

9. Department of Archaeology, National Museums of Kenya, Nairobi, Kenya

Angela Kabiru & Chapurukha M. Kusimba

10. British Institute of Eastern Africa, Nairobi, Kenya

Angela Kabiru

11. Department of Archaeology and Heritage Studies, University of Dar es Salaam, Dar es Salaam, Tanzania

Audax Z. P. Mabulla

12. Department of Earth Sciences, National Museums of Kenya, Nairobi, Kenya

Fredrick K. Manthi, Emmanuel Ndiema & Christine Ogola

13. Cleveland Museum of Natural History, Cleveland, OH, USA

Elizabeth Sawchuk

14. Department of Anthropology, University of Alberta, Edmonton, Alberta, Canada

Elizabeth Sawchuk

15. Department of Anthropology, Stony Brook University, Stony Brook, NY, USA

Elizabeth Sawchuk

16. Department of Genetics and Genomics, College of Medicine and Health Sciences, United Arab Emirates University, Al-Ain, United Arab Emirates

Lihadh Al-Gazali, Bassam R. Ali & Salma Ben-Salem

17. Laboratoire Evolution et Santé Orale, Faculté de Chirurgie Dentaire, Université Toulouse III—Paul Sabatier, Toulouse, France

Thierry Letellier & Denis Pierron

18. Institut de Civilisations/Musée d’Art et d’Archéologie, Université d’Antananarivo, Antananarivo, Madagascar

Chantal Radimilahy & Jean-Aimé Rakotoarisoa

19. Department of Anthropology, Lehman College and The Graduate Center, The City University of New York, New York, NY, USA

Ryan L. Raaum

20. The New York Consortium in Evolutionary Primatology, New York, NY, USA

Ryan L. Raaum

21. Institutes of Energy and the Environment, The Pennsylvania State University, University Park, PA, USA

Brendan J. Culleton

22. Broad Institute of Harvard and MIT, Cambridge, MA, USA

Swapan Mallick, Nick Patterson & David Reich

23. National Museums of Kenya, Lamu Museums, Lamu, Kenya

Mohammed Ali Mwenje

24. Coastal Archaeology, Gede National Monument, Gede, Kenya

Khalfan Bini Ahmed

25. Coastal Archaeology, Fort Jesus Museum, Mombasa, Kenya

Mohamed Mchulla Mohamed

26. Department of Anthropology, University of Illinois at Chicago, Chicago, IL, USA

Sloan R. Williams

27. University of Pennsylvania Museum of Archaeology and Anthropology, Philadelphia, PA, USA

Janet Monge

28. Department of Anthropology, University of South Florida, Tampa, FL, USA

Sibel Kusimba & Chapurukha M. Kusimba

29. Institute of African Studies, University of Nairobi, Museum Hill, Nairobi, Kenya

Chapurukha M. Kusimba

Contributions

E.S.B., J.F., S.W.-J., S.R.W., J.M., M.E.P., D.R. and C.M.K. conceptualized the study. E.S.B., K.S. and S.M. formally analysed the genetic data. E.S.B., N.B., K.C., E.C., L.I., A.M.L., J.O., L.Q., K.S., J.N.W., F.Z., B.J.C., S.M., N.R., N.P., K.B.A., M.M.M., S.K. and D.R. were involved in the genetic data investigation. J.F., S.W.-J., J.O., G.A., A.O.G., A. Kabiru, A. Kwekason, A.Z.P.M., F.K.M., E.N., C.O., E.S., L.A.-G., B.R.A., S.B.-S., T.L., D.P., C.R., J.-A.R., R.L.R., B.J.C., M.A.M., K.B.A., M.M.M., S.R.W., J.M., S.K., M.E.P., D.R. and C.M.K. provided resources. E.S.B., J.F., S.W.-J., M.E.P., D.R. and C.M.K. wrote the paper.

Corresponding authors

Correspondence to Esther S. Brielle, Jeffrey Fleisher, Stephanie Wynne-Jones, David Reich or Chapurukha M. Kusimba.

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks Annalisa Christie, Diyendo Massilani and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer review reports are available.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 ADMIXTURE analysis.

We analyzed 6422 individuals including African populations genotyped with the Human Origins SNP array, along with 1240K capture ancient individuals from present-day Kenya, present-day Tanzania, and surrounding areas; Swahili coast present-day samples;[21,32] ancient Israel_Natufian; and Pulliyar samples from present-day India. We chose K = 10 ancestral reference populations (chosen based on low cross-validation errors, high log-likelihood scores, and a low number of reference populations so as to not overfit) to highlight various ancestries found throughout Eastern Africa). We were able to discern a variety of African pastoral, farming, and forager ancestries within coastal and inland Swahili-speaking populations of the Human Origins dataset. The Makwasinyi reference ancestry composition resemble those of the present-day coastal populations sampled in a previously published dataset[32], albeit with different proportions. The x's identify two outlier individuals from the dataset with more Asian admixture; we removed them for population-level modeling. The genome-wide ancestry composition of the ancient coastal samples more closely resembles that of the newly published present-day individuals for which haplogroup data was published in[21].

Extended Data Fig. 2 Relationship of medieval to recent East Africans.

A) Graphical representation of individual and population level *qpAdm* models for Makwasinyi individuals, 89 present-day people who identify as and speak Swahili[32], and 93 present-day people who identify as Swahili and have long familial residency in the town, sampled from 6 coastal regions along Kenya[21]. The green ancestry component represents the modeled contribution from medieval Swahili coast people, as proxied by individuals buried in Kilwa or Manda. One standard error bars around the mean are computed using a Block Jackknife across all 5 centimorgan segments of the autosomes, and are meaningful even for single individuals as the genome reflects ancestral contributions of large numbers of ancestors. B) PCA with

the same present-day Swahili-identifying people (asterisks) and other present and ancient people (filled circles)[23,54,55], and ancient individuals from eastern Africa published in this study (open shapes)[23,24] projected onto the first two eigenvectors.

Extended Data Fig. 3 Estimates of dates of population mixture.

Curves that show the exponential decay of linkage disequilibrium generated by mixture between two populations related differentially to the two sources as a function of the number of elapsed generations since the mixture event. Multiplying by a 28 ± 2 year-to-generation conversion factor[53] and subtracting from the average calibrated radiocarbon dates or the archaeologically estimated date of the ancient individuals gives estimates in calendar years. For present-day Swahili individuals, we subtract from the year 2015, when sampling occured.

Extended Data Table 1 Ancient individuals with high coverage genome-wide data newly reported in this study
Extended Data Table 2 Ancient individuals with high coverage genome-wide data newly reported in this study that have significant evidence of Persian and often Indian ancestry
Extended Data Table 3 Individuals with high coverage genome-wide data that do not have evidence of Persian ancestry
Extended Data Table 4 Proportions of ancestry for present-day Swahili-identified individuals newly genotyped for this study

Supplementary information

Supplementary Information

Reporting Summary

Peer Review File

Supplementary Data Files

This file contains Supplementary Data files 1–7.

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Brielle, E.S., Fleisher, J., Wynne-Jones, S. *et al.* Entwined African and Asian genetic roots of medieval peoples of the Swahili coast. *Nature* **615**, 866–873 (2023). https://doi.org/10.1038/s41586-023-05754-w

- Received: 07 July 2022
- Accepted: 24 January 2023
- Published: 29 March 2023
- Issue Date: 30 March 2023
- DOI: https://doi.org/10.1038/s41586-023-05754-w

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

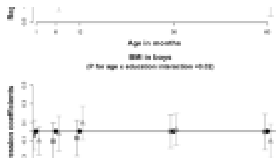
Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

Medieval Swahili people had both African and Asian ancestry

Nature Research Briefing 29 Mar 2023

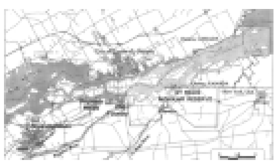
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05754-w



Socioeconomic inequalities in weight, height and body mass index from birth to 5 years

17 August 2018

M. Ballon, J. Botton, … on behalf of the EDEN Mother–Child Cohort Study Group



Trends in height, weight, BMI, skinfolds, and measures of overweight and obesity from 1979 through 1999 among American Indian Youth: The Akwesasne Mohawk

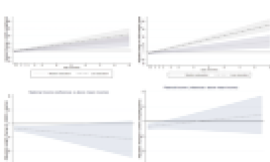
03 April 2019

Lawrence M. Schell, Mia V. Gallo, … the Akwesasne Task Force on the Environment

The effect of height on estimates of the change in BMI-based prevalence of childhood obesity

18 August 2021

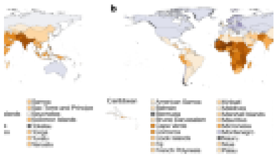
Richard D. Telford, Timothy S. Olds, … Ross B. Cunningham



[Socioeconomic inequalities in children’s weight, height and BMI trajectories in Norway]

02 March 2021

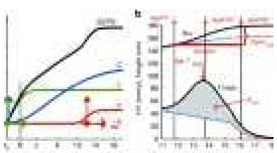
Teferi Mekonnen, Eleni Papadopoulou, … Mekdes K. Gebremariam



[Rising rural body-mass index is the main driver of the global obesity epidemic in adults]

08 May 2019

NCD Risk Factor Collaboration (NCD-RisC)



[Estimating secular changes in longitudinal growth patterns underlying adult height with the QEPS model: the Grow Up Gothenburg cohorts]

23 May 2018

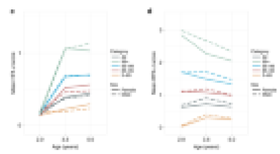
Anton Holmgren, Aimon Niklasson, … Kerstin Albertsson-Wikland



[Trajectory of body mass index and height changes from childhood to adolescence: a nationwide birth cohort in Japan]

26 November 2021

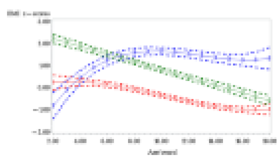
Naomi Matsumoto, Toshihide Kubo, … Takashi Yorifuji



[Longitudinal trajectories of adiposity-related measures from age 2–5 years in a population of low-income Hispanic children]

04 August 2020

Catherine M. Crespi, Shuang Gao, … May C. Wang



[Associations of community, famliy and early individual factors with body mass index z-scores trajectories among Chinese children and adolescents]

15 July 2021

Jing Liang, Si Zheng, … Peigang Wang

- Article
- Open Access
- Published: 29 March 2023

Diminishing benefits of urban living for children and adolescents' growth and development

- NCD Risk Factor Collaboration (NCD-RisC)

Nature **volume 615**, pages 874–883 (2023)

- 5128 Accesses
- 206 Altmetric
- Metrics details

Abstract

Optimal growth and development in childhood and adolescence is crucial for lifelong health and well-being[1,2,3,4,5,6]. Here we used data from 2,325 population-based studies, with measurements of height and weight from 71 million participants, to report the height and body-mass index (BMI) of children and adolescents aged 5–19 years on the basis of rural and urban place of residence in 200 countries and territories from 1990 to 2020. In 1990, children and adolescents residing in cities were taller than their rural counterparts in all but a few high-income countries. By 2020, the urban height advantage became smaller in most countries, and in many high-income western countries it reversed into a small urban-based disadvantage. The exception was for boys in most countries in sub-Saharan Africa and in some countries in Oceania, south Asia and the region of central Asia, Middle East and north Africa. In these countries, successive cohorts of boys from rural places either did not gain height or possibly became shorter, and hence fell further behind their urban peers. The difference between the age-standardized mean BMI of children in urban and rural areas was <1.1 kg m^{-2} in the vast majority of countries. Within this small range, BMI increased slightly more in cities than in rural areas, except in south Asia, sub-Saharan Africa and some countries in central and eastern Europe. Our results show that in much of the world, the growth and developmental advantages of living in cities have diminished in the twenty-first century, whereas in much of sub-Saharan Africa they have amplified.

Main

The growth and development of school-aged children and adolescents (ages 5–19 years) are influenced by their nutrition and environment at home, in the community and at school. Healthy growth and development at these ages help consolidate gains and mitigate inadequacies from early childhood and vice versa[1], with lifelong implications for health and well-being[2,3,4,5,6]. Until recently, the growth and development of older children and adolescents received substantially less attention than in early childhood and adulthood[7]. Increasing attention on the importance of health and nutrition during school years has been accompanied by a presumption that differences in nutrition and the environment lead to distinct, and generally less healthy, patterns of growth and development at these ages in cities compared to rural areas[8,9,10,11,12,13,14,15,16,17]. This presumption is despite some empirical studies showing that food quality and nutrition are better in cities[18,19].

Data on growth and developmental outcomes during school ages are needed, alongside data on the efficacy of specific interventions and policies, to select and prioritize policies and programmes that promote health and health equity, both for the increasing urban population and for children who continue to grow up in rural areas. Consistent and comparable global data also help benchmark across countries and territories and draw lessons on good practice. Yet, globally, there are fewer data on growth trajectories in rural and urban areas in these formative ages than for children under 5 years of age[20] or for adults[21]. The available studies have been in one country, at one point in time and/or in one sex and narrow age groups. The few studies that covered more than one country[22,23,24] mostly focused on older girls and used at most a few dozen data sources and hence could not systematically measure long-term trends. Consequently, many policies and programmes that aim to enhance healthy growth and development in school ages focus narrowly and generically on specific features of nutrition or the environment in either cities or rural areas[10,13,25,26,27,28]. Little attention has been paid to the similarities and differences between relevant outcomes in these settings or to the heterogeneity of the urban–rural differences across countries.

Here we report on the mean height and BMI of school-aged children and adolescents residing in rural and urban areas of 200 countries and territories (referred to as countries hereafter) from 1990 to 2020. Height and BMI are anthropometric measures of growth and development that are

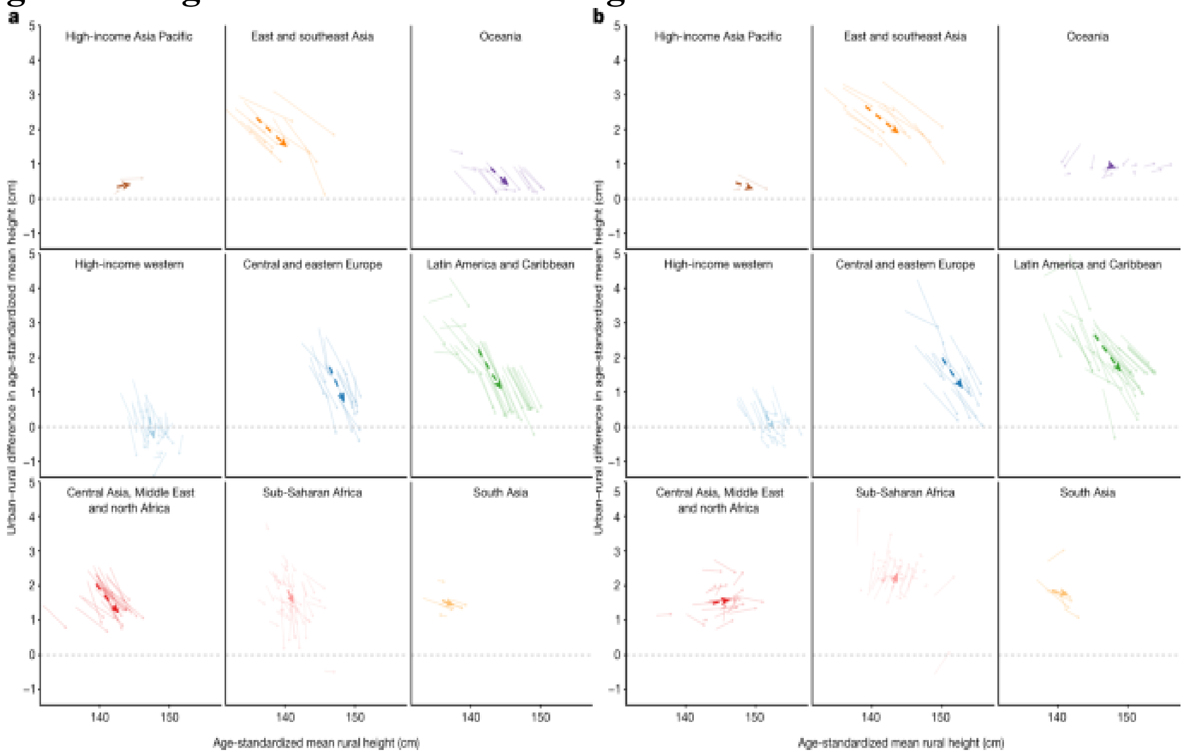
influenced by the quality of nutrition and healthiness of the living environment and are highly predictive of health and well-being throughout life in observational and Mendelian randomization studies[2,3,4,5,6]. These studies have shown that having low height and excessively low BMI increases the risk of morbidity and mortality, and low height impairs cognitive development and reduces educational performance and work productivity in later life[2,3,4]. A high BMI in these ages increases the lifelong risk of overweight and obesity and several non-communicable diseases, and might contribute to poor educational outcomes[5,6].

We used 2,325 population-based studies that measured height and weight in 71 million participants in 194 countries (Extended Data Fig. 1 and Supplementary Table 2). We used these data in a Bayesian hierarchical meta-regression model to estimate mean height and BMI of children and adolescents aged 5–19 years by rural and urban place of residence, year and age for 200 countries. Details of data sources and statistical methods are provided in the Methods. Our results represent the height and BMI for children and adolescents of the same age over time (that is, successive cohorts) in rural and urban areas of each country, and the difference between the two. For presentation, we summarize the 15 age-specific estimates, for single years of age from 5 to 19, through age standardization, which puts each country-year's child and adolescent population on the same age distribution and enables comparisons to be made over time and across countries. We also show results, graphically and numerically, for index ages of 5, 10, 15 and 19 years in the Supplementary Information.

In 1990, school-aged boys and girls who lived in cities had a height advantage (that is, were taller) compared with their rural counterparts. The exception was in high-income countries, where the urban height advantage was either negligible (<1.2 cm for age-standardized mean height; posterior probability (PP) for children living in urban areas being taller ranging from 0.51 to >0.99) or there was a small rural advantage (for example, Belgium, the Netherlands and the United Kingdom) (PP for children in rural areas being taller ranging from 0.53 to >0.95 where there was a rural height advantage) (Fig. 1 and Extended Data Fig. 2). The largest height differences between children and adolescents in cities and rural areas in 1990 occurred in some countries in Latin America (for example, Mexico, Guatemala,

Panama and Peru), east and southeast Asia (China, Indonesia and Vietnam), central and eastern Europe (Bulgaria, Hungary and Romania) and sub-Saharan Africa (Democratic Republic of Congo (DR Congo) and Rwanda). The urban height advantage in boys and girls in the named countries ranged from 2.4 to 5.0 cm, and the PP of children living in urban areas being taller than children living in rural areas was >0.99 (see Supplementary Table 3 for country-specific numerical values of height in children living in rural versus urban areas, their difference and the corresponding credible intervals (CrIs)).

Fig. 1: Change in the urban–rural height difference from 1990 to 2020.



a,**b**, Change in the urban–rural difference in age-standardized mean height in relation to the change in age-standardized mean rural height in girls (**a**) and boys (**b**). Each solid arrow in lighter shade shows one country beginning in 1990 and ending in 2020. The dashed arrows in darker shade show the regional averages, calculated as the unweighted arithmetic mean of the values for all countries in each region along the horizontal and vertical axes. For the urban–rural difference, a positive number shows a higher urban mean height and a negative number shows higher rural mean height. See Extended Data Fig. 2 for urban–rural differences in age-standardized mean height and their change over time shown as maps, together with

uncertainties in the estimates. See Supplementary Fig. 4a for results at ages 5, 10, 15 and 19 years. We did not estimate the difference between rural and urban height for countries classified as entirely urban (Bermuda, Kuwait, Nauru and Singapore) or entirely rural (Tokelau).

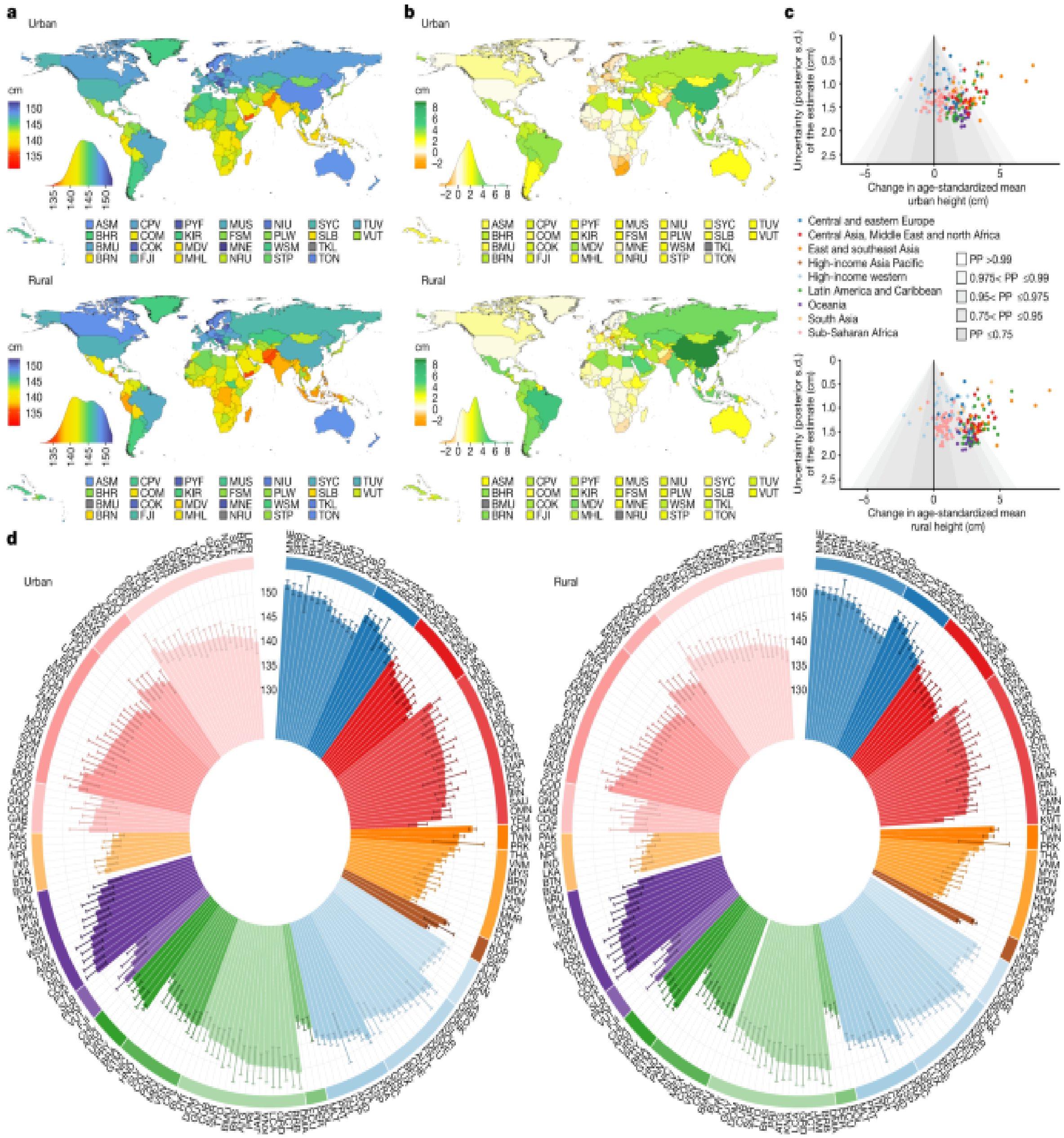
The urban–rural height gap in the late twentieth century among low-income and middle-income countries was determined by how much children and adolescents in cities and rural areas had approached as opposed to fallen behind their peers in high-income countries, where there was little difference between urban and rural height. In countries such as Bulgaria, Hungary and Romania, the height of children and adolescents living in urban areas approached that of high-income countries, whereas children and adolescents in rural areas lagged behind, leading to a relatively large gap. In much of sub-Saharan Africa and south Asia, the height of children and adolescents lagged behind their peers in high-income countries regardless of where they lived, such that the urban–rural gap was relatively small. In a third group of low-income or middle-income countries that included Indonesia, Vietnam, Panama, Peru, DR Congo and Rwanda, children living in urban areas remained shorter than in high-income countries, but children from rural areas lagged even further behind, such that the urban–rural gap became large.

By 2020, the urban height advantage in school ages became smaller in much of the world. In many high-income western countries and some central European countries, it disappeared or reversed into a small (typically <1 cm) urban disadvantage (Fig. 1 and Extended Data Figs. 2 and 8). Countries with substantial convergence over these three decades were in central and eastern Europe (for example, Croatia), Latin America and the Caribbean (for example, Argentina, Brazil, Chile and Paraguay), east and southeast Asia (for example, Taiwan) and for girls in central Asia (for example, Kazakhstan and Uzbekistan). The urban height advantage in the named countries declined by around 1–2.5 cm from 1990 to 2020 (the PP of urban–rural height difference having declined ≥0.90 for the named countries). In many other middle-income countries (for example, China, Romania and Vietnam), the urban–rural height gaps declined, but children and adolescents living in cities remained taller than their rural counterparts (by 1.7–2.5 cm in the named countries for boys and girls; the PP of children in cities being taller than children in rural areas in 2020 >0.99). The exception to this

convergence was for boys in most countries in sub-Saharan Africa and some countries in Oceania, south Asia and the region of central Asia, Middle East and north Africa, where the urban height advantage slightly increased over these three decades. The largest increase in the urban height advantage for boys occurred in countries in east Africa such as Ethiopia (0.9 cm larger height gap in 2020 than 1990; 95% CrI −0.9 to 2.9, and PP of an increase of 0.86), Rwanda (1.0 cm larger gap, 95% CrI −0.7 to 3.0, and PP 0.88) and Uganda (1.1 cm larger gap, 95% CrI of −0.6 to 3.1, and PP 0.89). For girls, the urban–rural gap remained largely unchanged in many countries in sub-Saharan Africa and south Asia.

In middle-income countries and emerging economies (newly high-income and industrialized countries) where the height of children and adolescents residing in rural areas converged to those in cities, successive cohorts of children and adolescents living in rural areas outpaced their urban counterparts in becoming taller and attained heights that urban children in the same countries had done decades earlier: growing to heights closer to those seen in high-income countries (Figs. 2 and 3). Successive cohorts of children and adolescents residing in rural areas in sub-Saharan Africa did not experience the accelerated height gain seen in cohorts in rural areas of middle-income countries. Notably, in the case of boys living in sub-Saharan Africa, there was no gain, or possibly a decrease, in height, which in turn led to a persistence or even widening of the urban–rural gap. As a result of these global trends, by 2020, the largest urban–rural gaps in height were seen in Andean and central Latin America (for example, Bolivia, Panama and Peru, by up to 4.7 cm (95% CrI 4.0–5.5 cm) for boys and 3.8 cm (95% CrI 3.3–4.3 cm) for girls) and, especially for boys, in sub-Saharan Africa (for example, DR Congo, Ethiopia, Mozambique and Rwanda, by up to 4.2 cm (95% CrI 2.7–5.7 cm)).

Fig. 2: Urban and rural height in 2020 and the change from 1990 to 2020 for girls.

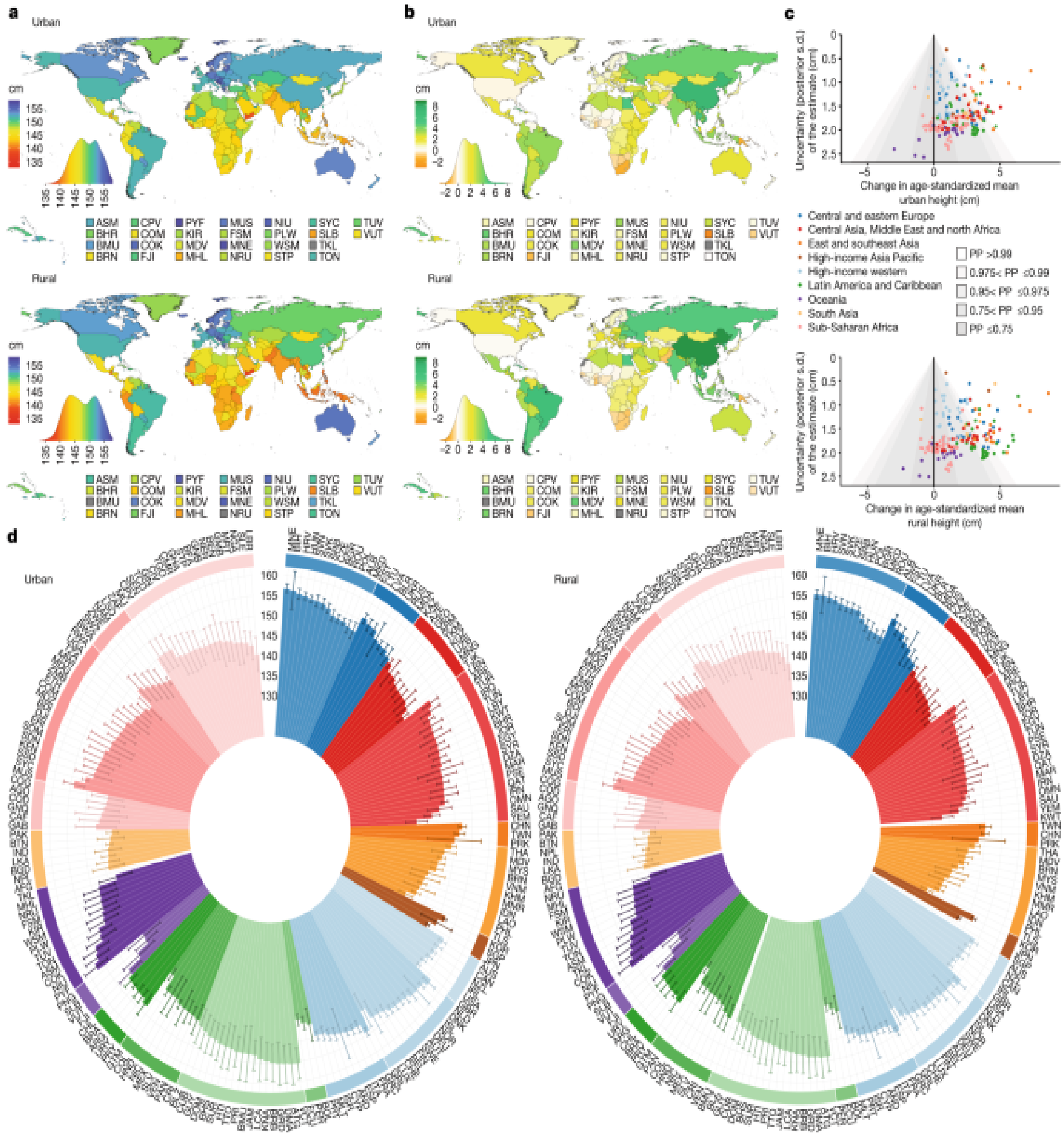


a, Age-standardized mean height in 2020 by urban and rural place of residence for girls. The density plots show the distribution of estimates across countries. **b**, Age-standardized change in mean height from 1990 to 2020 by urban and rural place of residence for girls. The density plots show the distribution of estimates across countries. **c**, Change in mean height from 1990 to 2020 in relation to the uncertainty of the change measured by posterior standard deviation. Each point in the scatter plots shows one country. Shaded areas approximately show the PP of an estimated change being a true increase or decrease. The PP of a decrease is one minus that of

an increase. If an increase in mean height is statistically indistinguishable from a decrease, the PP of an increase and a decrease is 0.50. PPs closer to 0.50 indicate more uncertainty, whereas those towards 1 indicate more certainty of change. **d**, Age-standardized mean height in 2020 for all countries. The height of each column is the posterior mean estimate shown together with its 95% CrI. Countries are ordered by region and super-region. See Extended Data Fig. 4 for a map of PP of the estimated change. See Supplementary Fig. 5 for results at ages 5, 10, 15 and 19 years. See Supplementary Table 3 for numerical results, including Crls, as age-standardized and at ages 5, 10, 15 and 19 years. We did not estimate mean rural height in countries classified as entirely urban (Bermuda, Kuwait, Nauru and Singapore), mean urban height in countries classified as entirely rural (Tokelau) or their change over time in these countries, as indicated in grey. Countries are labelled using their International Organization for Standardization (ISO) 3166-1 alpha-3 codes. Afghanistan, AFG; Albania, ALB; Algeria, DZA; American Samoa, ASM; Andorra, AND; Angola, AGO; Antigua and Barbuda, ATG; Argentina, ARG; Armenia, ARM; Australia, AUS; Austria, AUT; Azerbaijan, AZE; Bahamas, BHS; Bahrain, BHR; Bangladesh, BGD; Barbados, BRB; Belarus, BLR; Belgium, BEL; Belize, BLZ; Benin, BEN; Bermuda, BMU; Bhutan, BTN; Bolivia, BOL; Bosnia and Herzegovina, BIH; Botswana, BWA; Brazil, BRA; Brunei Darussalam, BRN; Bulgaria, BGR; Burkina Faso, BFA; Burundi, BDI; Cabo Verde, CPV; Cambodia, KHM; Cameroon, CMR; Canada, CAN; Central African Republic, CAF; Chad, TCD; Chile, CHL; China, CHN; Colombia, COL; Comoros, COM; Congo, COG; Cook Islands, COK; Costa Rica, CRI; Cote d'Ivoire, CIV; Croatia, HRV; Cuba, CUB; Cyprus, CYP; Czechia, CZE; Denmark, DNK; Djibouti, DJI; Dominica, DMA; Dominican Republic, DOM; DR Congo, COD; Ecuador, ECU; Egypt, EGY; El Salvador, SLV; Equatorial Guinea, GNQ; Eritrea, ERI; Estonia, EST; Eswatini, SWZ; Ethiopia, ETH; Fiji, FJI; Finland, FIN; France, FRA; French Polynesia, PYF; Gabon, GAB; Gambia, GMB; Georgia, GEO; Germany, DEU; Ghana, GHA; Greece, GRC; Greenland, GRL; Grenada, GRD; Guatemala, GTM; Guinea Bissau, GNB; Guinea, GIN; Guyana, GUY; Haiti, HTI; Honduras, HND; Hungary, HUN; Iceland, ISL; India, IND; Indonesia, IDN; Iran, IRN; Iraq, IRQ; Ireland, IRL; Israel, ISR; Italy, ITA; Jamaica, JAM; Japan, JPN; Jordan, JOR; Kazakhstan, KAZ; Kenya, KEN; Kiribati, KIR; Kuwait, KWT; Kyrgyzstan, KGZ; Lao PDR, LAO; Latvia, LVA; Lebanon, LBN; Lesotho,

LSO; Liberia, LBR; Libya, LBY; Lithuania, LTU; Luxembourg, LUX; Madagascar, MDG; Malawi, MWI; Malaysia, MYS; Maldives, MDV; Mali, MLI; Malta, MLT; Marshall Islands, MHL; Mauritania, MRT; Mauritius, MUS; Mexico, MEX; Micronesia (Federated States of), FSM; Moldova, MDA; Mongolia, MNG; Montenegro, MNE; Morocco, MAR; Mozambique, MOZ; Myanmar, MMR; Namibia, NAM; Nauru, NRU; Nepal, NPL; Netherlands, NLD; New Zealand, NZL; Nicaragua, NIC; Niger, NER; Nigeria, NGA; Niue, NIU; North Korea, PRK; North Macedonia, MKD; Norway, NOR; Occupied Palestinian Territory, PSE; Oman, OMN; Pakistan, PAK; Palau, PLW; Panama, PAN; Papua New Guinea, PNG; Paraguay, PRY; Peru, PER; Philippines, PHL; Poland, POL; Portugal, PRT; Puerto Rico, PRI; Qatar, QAT; Romania, ROU; Russian Federation, RUS; Rwanda, RWA; Saint Kitts and Nevis, KNA; Saint Lucia, LCA; Samoa, WSM; Sao Tome and Principe, STP; Saudi Arabia, SAU; Senegal, SEN; Serbia, SRB; Seychelles, SYC; Sierra Leone, SLE; Singapore, SGP; Slovakia, SVK; Slovenia, SVN; Solomon Islands, SLB; Somalia, SOM; South Africa, ZAF; South Korea, KOR; South Sudan, SSD; Spain, ESP; Sri Lanka, LKA; Saint Vincent and the Grenadines, VCT; Sudan, SDN; Suriname, SUR; Sweden, SWE; Switzerland, CHE; Syrian Arab Republic, SYR; Taiwan, TWN; Tajikistan, TJK; Tanzania, TZA; Thailand, THA; Timor-Leste, TLS; Togo, TGO; Tokelau, TKL; Tonga, TON; Trinidad and Tobago, TTO; Tunisia, TUN; Turkey, TUR; Turkmenistan, TKM; Tuvalu, TUV; Uganda, UGA; Ukraine, UKR; United Arab Emirates, ARE; United Kingdom, GBR; United States of America, USA; Uruguay, URY; Uzbekistan, UZB; Vanuatu, VUT; Venezuela, VEN; Vietnam, VNM; Yemen, YEM; Zambia, ZMB.

Fig. 3: Urban and rural height in 2020 and change from 1990 to 2020 for boys.

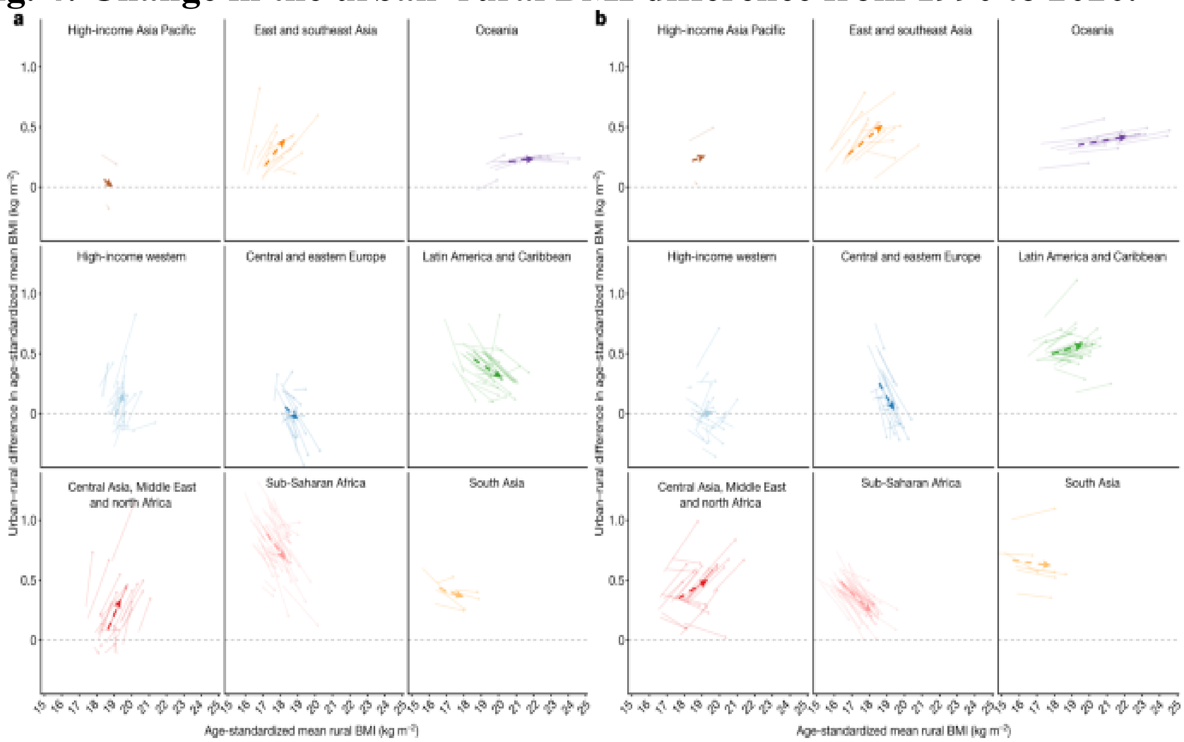


a–**d**, See the caption for Fig. 2 for descriptions of the contents of the figure and for definitions. We did not estimate mean rural height in countries classified as entirely urban (Bermuda, Kuwait, Nauru and Singapore), mean urban height in countries classified as entirely rural (Tokelau) or their change over time, as indicated in grey.

The urban–rural BMI difference was relatively small throughout these three decades, <1.4 kg m^{-2} in all countries and years and <1.1 kg m^{-2} in all but nine countries, for age-standardized mean BMI (Fig. 4 and Extended Data

Figs. 3 and 9). In 1990, the urban–rural BMI gap was largest in sub-Saharan Africa (for example, Ethiopia, Kenya, Malawi, South Africa and Zimbabwe) and south Asia (for example, Bangladesh and India), followed by parts of Latin America (for example, Mexico and Peru). The urban–rural BMI gap in the two sexes in the named countries ranged from 0.4 to 1.2 kg m^{-2}, and the PP of children and adolescents living in urban areas having a higher BMI than those in rural areas was ≥0.89. At that time, girls and/or boys in rural areas of some of these countries had mean BMI levels that were close to, and in some ages below, the thresholds of being underweight (>1 s.d. below the median of the World Health Organization (WHO) reference population).

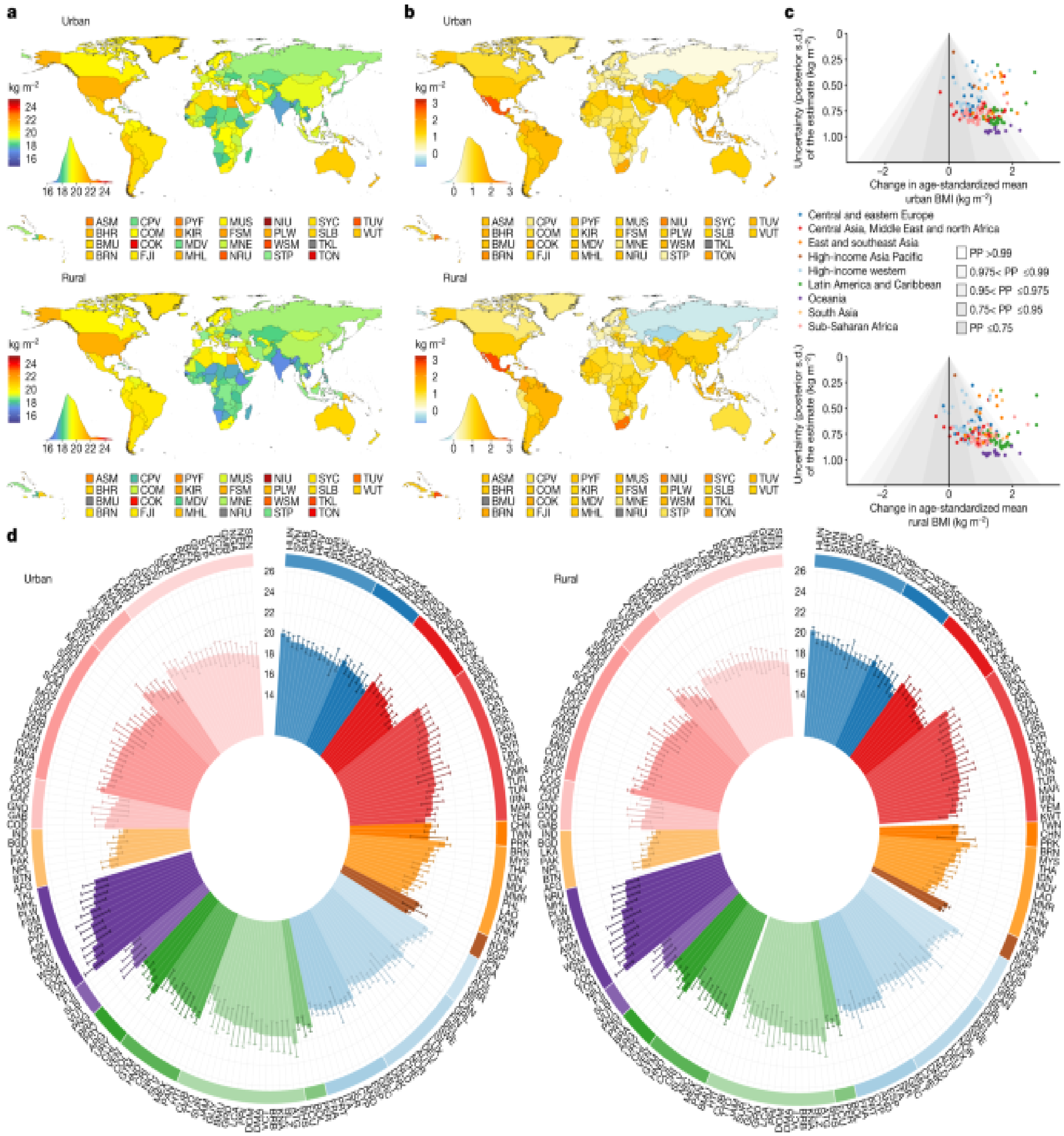
Fig. 4: Change in the urban–rural BMI difference from 1990 to 2020.



a,**b**, Change in urban–rural difference in age-standardized mean BMI for girls (**a**) and boys (**b**) in relation to change in age-standardized mean rural BMI. See the caption for Fig. 1 for a description of the contents of this figure. See Extended Data Fig. 3 for urban–rural differences in age-standardized mean BMI and their change over time shown as maps, together with uncertainties in the estimates. See Supplementary Fig. 4b for results at ages 5, 10, 15 and 19 years. We did not estimate the difference between rural and urban BMI for countries classified as entirely urban (Bermuda, Kuwait, Nauru and Singapore) or entirely rural (Tokelau).

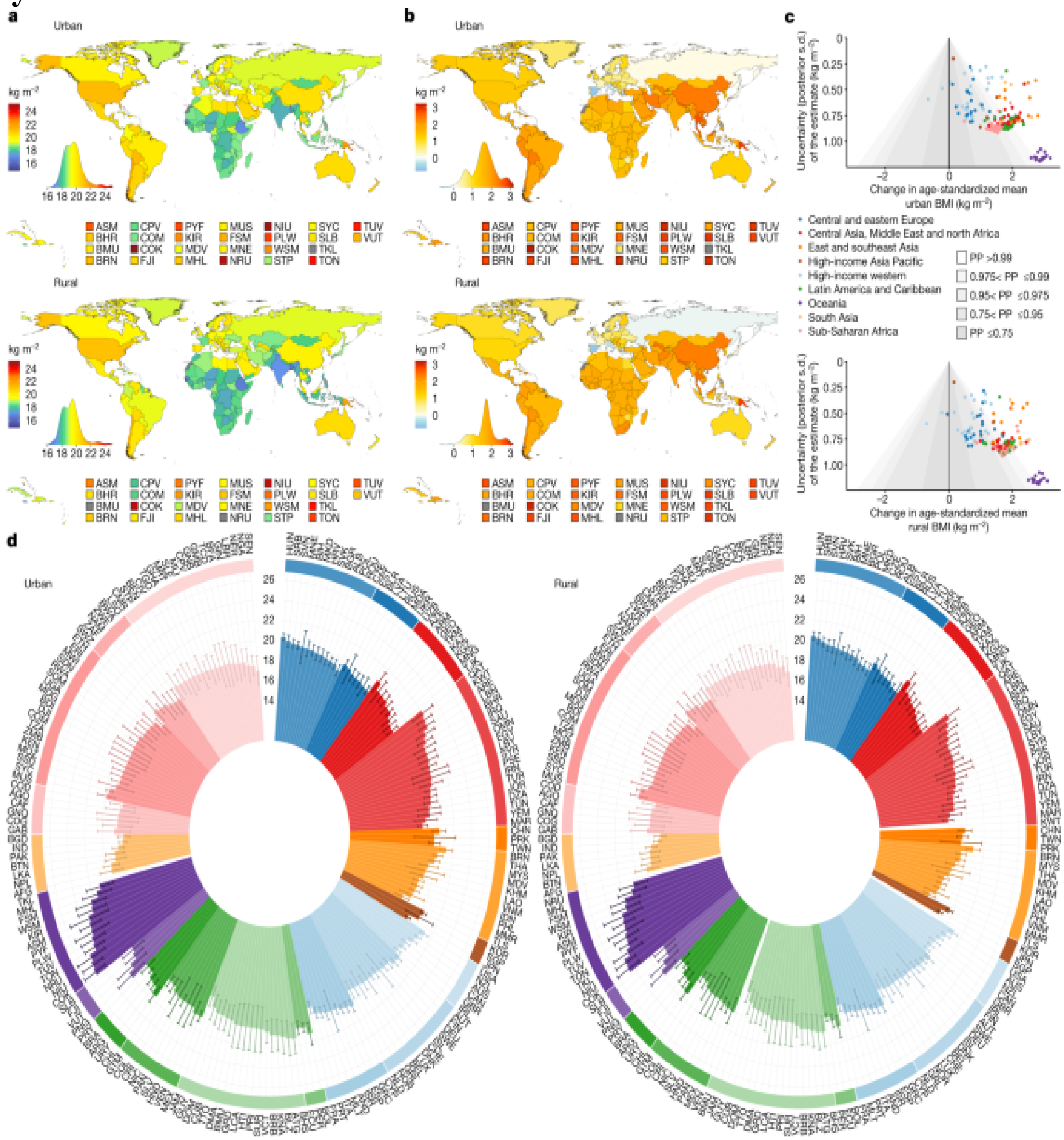
From 1990 to 2020, the BMI of successive cohorts of children and adolescents in both urban and rural areas increased in all but a few mostly high-income countries (for example, Denmark, Italy and Spain) (Figs. 5 and 6). There was heterogeneity in low-income and middle-income countries in how much the BMI increased in cities compared with rural areas. In the majority of countries in sub-Saharan Africa and south Asia, the BMI of successive cohorts of children and adolescents increased more in rural areas than in cities, leading to a closing of the urban–rural difference. The urban–rural BMI gap declined by up to 0.65 kg m^{-2} for both girls and boys, and the PP that the urban–rural BMI difference declined from 1990 to 2020 ranged from 0.52 to 0.95. In both sub-Saharan Africa and south Asia, these changes shifted the mean BMI of boys and girls in rural areas out of the range for being underweight. Moreover, in many countries in sub-Saharan Africa, this shift continued beyond the median of the WHO reference population and in some cases approached the threshold for being overweight (>1 s.d. above the median of the WHO reference population). The opposite, a larger increase in urban BMI, happened in most other low-income and middle-income countries, leading to a slightly larger urban BMI excess in 2020 than in 1990. High-income countries and those in central and eastern Europe experienced a mix of increasing and decreasing urban BMI excess, but remained within a small range (−0.3 to 0.6 kg m^{-2} for almost all countries) over the entire period of analysis. At the regional level, the urban–rural BMI difference changed by <0.25 kg m^{-2} in these regions.

Fig. 5: Urban and rural BMI in 2020 and change from 1990 to 2020 for girls.



a–**d**, See the caption for Fig. 2 for descriptions of the contents of the figure and for definitions. See Extended Data Fig. 5 for a map of PP of the estimated change. See Supplementary Fig. 6 for results at ages 5, 10, 15 and 19 years. See Supplementary Table 4 for numerical results, including CrIs, as age-standardized and at ages 5, 10, 15 and 19 years. We did not estimate mean rural BMI in countries classified as entirely urban (Singapore, Bermuda and Nauru), mean urban BMI in areas classified as entirely countries (Tokelau) or their change over time, as indicated in grey.

Fig. 6: Urban and rural BMI in 2020 and change from 1990 to 2020 for boys.

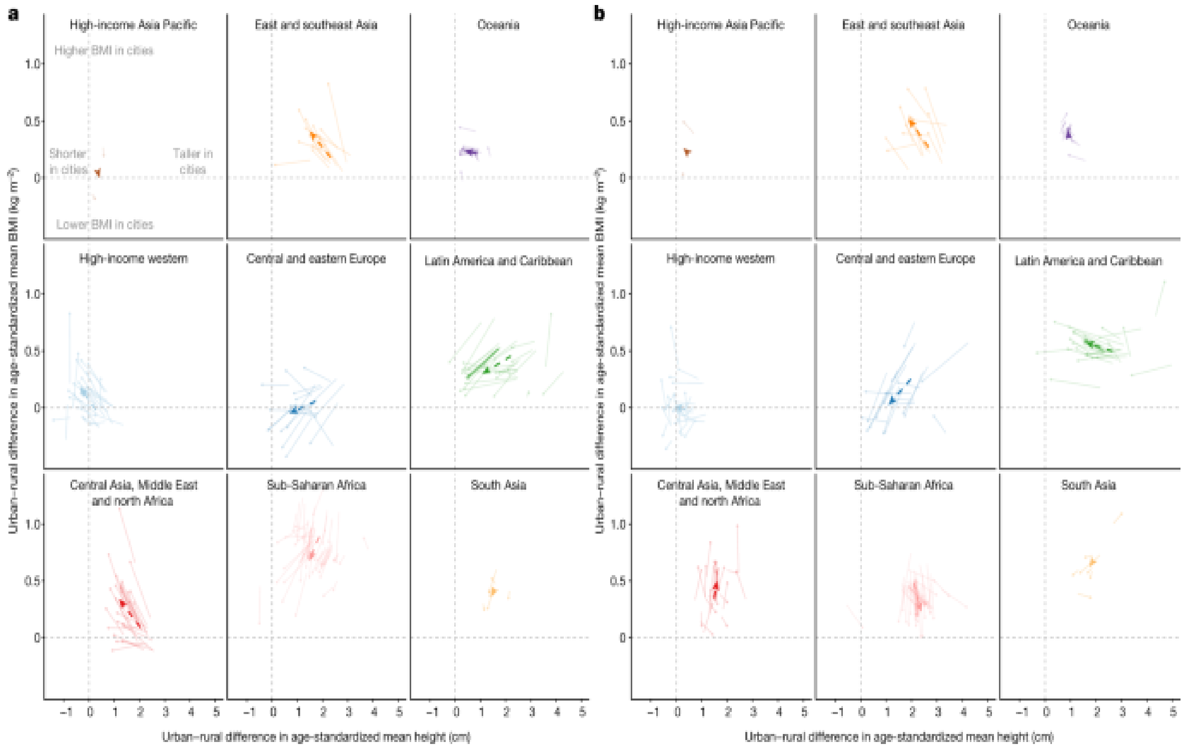


a–**d**, See the caption for Fig. 2 for descriptions of the contents of the figure and for definitions. We did not estimate mean rural BMI in countries classified as entirely urban (Singapore, Bermuda and Nauru), mean urban BMI in countries classified as entirely rural (Tokelau) or their change over time, as indicated in grey.

The urban height advantage was larger in boys than girls in most countries (Supplementary Fig. 3). Urban excess BMI was larger in boys than girls in only about one-half of the countries. For the other half, mostly in high-income western countries and those in sub-Saharan Africa, urban excess BMI was larger in girls than boys. The urban height advantage was slightly larger at 5 years of age than at 19 years of age in most low-income and middle-income countries, especially for girls, but there was little difference across ages in high-income regions and in central and eastern Europe (Supplementary Fig. 4).

Since the introduction of modern sanitation in the nineteenth century, cities provided substantial nutritional and health advantages in high-income and subsequently low-income and middle-income countries[19]. Our results show that in the twenty-first century, during school ages, these advantages have disappeared in high-income countries and diminished in middle-income countries and emerging economies in Asia, Latin America and the Caribbean, and parts of Middle East and north Africa. Specifically, in these settings, successive cohorts of school-aged children and adolescents living in cities were outpaced by those in rural areas in terms of height gain but gained slightly more weight by 2020, typically in the unhealthy range (Fig. 7). This contrasted with the poorest region in the world: sub-Saharan Africa. In this region, the urban height advantage persisted or even expanded, whereas rural mean BMI went beyond remedying underweight and surpassed the median of the WHO reference population in 2020, hence consolidating the urban advantage. South Asia had a mixed pattern of urban versus rural trends from 1990 to 2020, with children and adolescents in rural areas gaining both more height and more weight for their height than those in cities. Notably, our results also show that differences in height and BMI between urban and rural populations within most countries are smaller than the differences across countries, even those in the same region.

Fig. 7: Change in the urban–rural height and BMI difference from 1990 to 2020.



a,**b**, Change in the urban–rural difference in age-standardized mean height and the urban–rural difference in age-standardized mean BMI in girls (**a**) and boys (**b**). See the caption for Fig. 1 for a description of the contents of this figure. See Supplementary Fig. 4c for results at ages 5, 10, 15 and 19 years. We did not estimate the difference between rural and urban height and BMI for countries classified as entirely urban (Bermuda, Kuwait, Nauru and Singapore) or entirely rural (Tokelau).

We also found that the urban–rural BMI gap, although dynamic, changed much less than the BMI of either subgroup of the population and less than commonly assumed when discussing the role of cities in the obesity epidemic[8,10,12,13,15,16]. Urban–rural BMI differences were especially small in high-income countries, which is consistent with evidence from a few countries that show diets and behaviours are affected more by household socioeconomic status than whether children and adolescents live in cities or rural areas[29,30]. Urban BMI excess increased slightly more in middle-income countries in east and southeast Asia, Latin America and the Caribbean, and Middle East and north Africa, a trend that was the opposite of the convergence in BMI of adults in these same regions[21]. Additional analyses of data collected by the NCD Risk Factor Collaboration (NCD-RisC) for

young adults (20–29 and 30–39 years) showed that the shift from a small divergent trend to convergence of BMI between urban and rural areas happens in young adulthood (Extended Data Figs. 6 and 7), a period during which there is substantial, but variable, weight gain among population subgroups[31]. These shifts in trends from adolescence to young adulthood might be a result of changes in diet and energy expenditure that accompany changes in household structure, social and economic roles and the living environment[32,33,34].

Long-term follow-up studies have shown that children and adolescents do not achieve their height potential if they do not consume sufficient and diverse nutritious foods or if they are exposed to repeated or persistent infections, which result in loss of nutrients[2]. Studies that use data on household socioeconomic and environmental factors have indicated that these physiological determinants of height are themselves affected by income, education, quality of the living environment and access to healthcare in rural as well as urban areas[35]. This evidence indicates that the relatively small urban–rural height differentials in high-income countries may be because of a greater abundance of nutritious foods, including some fortified foods, better education and healthcare and greater ability to finance programmes that promote healthy growth in countries with greater per-capita income and better infrastructure. Variations across these countries in the urban–rural height gap within this small range may be due to the extent of socioeconomic inequalities and poverty, differences in the availability and cost of nutritious foods between cities and rural areas and whether there are specific programmes (for example, food assistance or school food programmes) that improve nutrition in disadvantaged groups[30,36,37]. The more marked changes in height in urban versus rural areas took place in middle-income countries and emerging economies. Case studies in some countries where the heights of children and adolescents living in rural and urban areas converged show that the convergence was partly due to using the growth in national income towards programmes and services that helped close gaps in nutrition, sanitation and healthcare between different areas and social groups[38,39,40]. In countries in central and eastern Europe, transition to a market economy and increases in trade may have reduced the disparity in access to, and seasonality of, healthy foods between urban and rural areas[41], and partly underlie the convergence of height seen in our results. By

contrast, case studies in some countries have shown that where economic growth was accompanied by large inequalities in income, nutrition and/or services, the urban advantage persisted[42,43,44].

The notable exception in the global trends was sub-Saharan Africa, where a stagnation or reversal of height gain in rural areas led to the persistence or widening of urban–rural height differences, whereas the opposite happened for BMI (Fig. 7). Case studies of specific countries have indicated that unfavourable trends in nutrition in rural Africa, where the majority of the poorest people in the world live, started from macroeconomic shocks in the late twentieth century[45] and subsequent agriculture, trade and development policies that limited improvements in income and services, and emphasized agricultural exports over local food security and diversity[45]. These macroeconomic factors in turn led to less diverse diets, with higher caloric intake rather than a shift to protein-rich and nutrient-rich foods (for example, animal products, seafood, fruits and vegetables)[46,47,48]. Moreover, the slow expansion of infrastructure and services in rural areas restricted improvements in other determinants of healthy growth, such as clean water, sanitation and health care[49].

Several other factors may have had a secondary role in the observed trends in height and BMI and their difference in rural and urban areas. First, weight gain during childhood may reduce the age of puberty onset, which in turn may limit height gain during adolescence[50,51]. No comparable global data currently exist on age at menarche and timing of pubertal growth, even at the national level. Second, rural-to-urban migration and reclassification of previously rural areas to urban as they grow and industrialize may account for some of the observed population-level trends. However, migration tends to be less common in childhood and adolescence than in adulthood in most countries. Finally, improvements in survival among children aged under 5 years in rural areas, particularly low-birthweight children, may have influenced the height and weight of those who survive beyond 5 years of age. However, current data on changes in child survival in rural and urban areas in sub-Saharan Africa are limited and inconclusive in terms of whether mortality declined faster in rural or urban areas[52,53].

As attention in global health turns to children and adolescents, there is a need to consider and evaluate how growth and development in these formative ages may be affected both by social and economic policies that influence household income and poverty and by programmes that affect nutrition, health services, infrastructure and living environments in rural and urban areas. The need to identify, implement and evaluate policies and programmes that improve growth and development outcomes is particularly relevant as the increase in poverty and the cost of food, especially of nutrient-rich foods, as a result of the macroeconomic changes resulting from the COVID-19 pandemic and the war in Ukraine, may hinder further gains or even set back healthy growth and development in children and adolescents.

Methods

We estimated trends in mean height and BMI for children and adolescents aged 5–19 years from 1990 to 2020 by rural and urban place of residence for the 200 countries and territories listed in Supplementary Table 1. We pooled, in a Bayesian meta-regression, repeated cross-sectional population-based data on height and BMI. Our results represent estimates of height and BMI for children and adolescents of the same age over time (that is, for successive cohorts) in rural and urban settings for each country.

Data sources

We used a database on cardiometabolic risk factors collated by NCD-RisC. Data were obtained from publicly available multi-country and national measurement surveys, for example, Demographic and Health Surveys (DHS), WHO-STEPwise approach to Surveillance (STEPS) surveys, and those identified through the Inter-University Consortium for Political and Social Research, UK Data Service and European Health Interview & Health Examination Surveys Database. With the help of the WHO and its regional and country offices as well as the World Heart Federation, we identified and accessed population-based survey data from national health and statistical agencies. We searched and reviewed published studies as previously detailed[54] and invited eligible studies to join NCD-RisC, as we did with data holders from earlier pooled analyses of cardiometabolic risk

factors[55,56,57,58]. The NCD-RisC database is continuously updated through all the above routes and through periodic requests to NCD-RisC members to ask them to suggest additional sources in their countries.

We carefully checked that each data source met our inclusion criteria, as listed below. Potential duplicate data sources were first identified by comparing studies from the same country and year, followed by checking with NCD-RisC members that had provided data about whether the sources from the same country and year, with similar samples, were the same or distinct. If two sources were confirmed as duplicates, one was discarded. All NCD-RisC members were also periodically asked to review the list of sources from their country to verify that the included data met the inclusion criteria and were not duplicates.

For each data source, we recorded the study population, the sampling approach, the years of measurement and the measurement methods. Only data that were representative of the population were included. All data sources were assessed in terms of whether they covered the entire country, one or more subnational regions (that is one or more provinces or states, more than three cities, or more than five rural communities), or one or a small number of communities (limited geographical scope not meeting above national or subnational criteria), and whether participants in rural, urban or both areas were included. As stated in the sections on the statistical model, these study-level attributes were used in the Bayesian hierarchical model to estimate mean height and BMI by country, year, sex, age and place of residence using all available data while taking into account differences in the populations from which different studies had sampled. All submitted data were checked by at least two independent individuals. Questions and clarifications were discussed with NCD-RisC members and resolved before data were incorporated into the database.

Anonymized individual data from the studies in the NCD-RisC database were re-analysed according to a common protocol. We calculated the mean height and the mean BMI, and the associated standard errors, by sex, single year of age from 5 to 19 years and rural or urban place of residence. Additionally, for analysis of height, participants aged 20–30 years were included, assigned to their corresponding birth cohort, because mean height in these ages would be at least that when they were aged 19 years given that

the decline in height with age begins in the third and fourth decades of life[59]. All analyses incorporated sample weights and complex survey design, when applicable, in calculating summary statistics. For studies that had used simple random sampling, we calculated the mean as the average of all individuals within the group and the associated standard error (s.d. divided by the square root of sample size); for studies that had used multistage (stratified) sampling, we accounted for survey design features, including clusters, strata and sample weights, to weight each observation by the inverse sampling probability and estimated standard error through Taylor series linearization, as implemented in the R 'survey' package[60]. Computer code was provided to NCD-RisC members who requested assistance. For surveys without information on the place of residence, we calculated summary statistics stratified by age and sex for the entire sample, which represented the population-weighted sum of rural and urban means; data on the share of population in urban and rural areas were from the United Nations Population Division[61].

Additionally, summary statistics for nationally representative data from sources that were identified but not accessed using the above routes were extracted from published reports. Data were also extracted for two STEPS surveys that were not publicly available. We also included data from a previous global-data pooling study[58], when not accessed through the above routes.

Data inclusion and exclusion

Data sources were included in the NCD-RisC height and weight database if the following criteria were met: measured data on height and weight were available; study participants were 5 years of age or older; data were collected using a probabilistic sampling method with a defined sampling frame; data were from population samples at the national, subnational or community level as defined above; and data were from the countries and territories listed in Supplementary Table 1.

We excluded all data sources that were solely based on self-reported weight and height without a measurement component because these data are subject to biases that vary by geography, time, age, sex and socioeconomic

characteristics[62,63,64]. Owing to these variations, approaches to correcting self-reported data may leave residual bias. We also excluded data sources on population subgroups for which anthropometric status may differ systematically from the general population, including the following: studies that had included or excluded people based on their health status or cardiovascular risk; studies in which participants were only ethnic minorities; specific educational, occupational or socioeconomic subgroups (with the exception noted below); those recruited through health facilities (with the exception noted below); and females aged 15–19 years in surveys that sampled only ever-married women or measured height and weight only among mothers.

We used school-based data in countries and age–sex groups with school enrolment of 70% or higher. We used data for which the sampling frame was health insurance schemes in countries where at least 80% of the population were insured. Finally, we used data collected through general practice and primary care systems in high-income and central European countries with universal insurance because contact with the primary care systems tends to be as good as or better than response rates for population-based surveys.

We excluded participants whose age was <18 years and whose data were not reported by single year of age (<0.01% of all participants) because height and weight may have nonlinear age associations in these ages, especially during growth spurts. We excluded BMI data for females who were pregnant at the time of measurement (<0.01% of all participants). We excluded <0.2% of all participants who had recorded height: <60 cm or >180 cm for ages <10 years; <80 cm or >200 cm for ages 10–14 years; <100 cm or >250 cm for ages ≥15 years, or who had recorded weight: <5 kg or >90 kg for age <10 years; <8 kg or >150 kg for ages 10–14 years; <12 kg or >300 kg for ages ≥15 years, or who had recorded BMI: <6 kg m^{-2} or >40 kg m^{-2} for ages <10 years; <8 kg m^{-2} or >60 kg m^{-2} for ages 10–14 years; <10 kg m^{-2} or >80 kg m^{-2} for ages ≥15 years.

Conversion of BMI prevalence metrics to mean BMI

In 0.5% of our data points, mostly extracted from published reports or from a previous pooling analysis[58], the mean BMI was not reported but data were

available for the prevalence of one or more BMI categories, for example BMI ≥30 kg m^{-2}. To use these data, we used previously validated conversion regressions[65] to estimate the missing primary outcome from the available BMI prevalence metric or metrics. Additional details on regression model specifications along with the regression coefficients are reported at https://github.com/NCD-RisC/ncdrisc-methods/.

Statistical model overview

We used a Bayesian hierarchical meta-regression model to estimate the mean height and BMI by country, year, sex, age and place of residence using the aforementioned data. For presentation, we summarized the 15 age-specific estimates, for single years of age from 5 to 19 years, through age standardization, which puts the child and adolescent population for each country-year on the same age distribution, and hence enables comparisons to be made over time and across countries. We generated age-standardized estimates by taking weighted means of age-specific estimates using age weights from the WHO standard population[66]. We also show results, graphically and numerically, for index ages of 5, 10, 15 and 19 years in the Supplementary Information.

The statistical model is described in detail in statistical papers[67,68], related substantive papers[7,20,21,55,56,57,58,65,69] and in the section below on model specification. In summary, the model had a hierarchical structure in which estimates for each country and year were informed by its own data, if available, and by data from other years in the same country and from other countries, especially those in the same region and super-region, with data for similar time periods. The extent to which estimates for each country-year were influenced by data from other years and other countries depended on whether the country had data, the sample size of the data, whether they were national, and the within-country and within-region variability of the available data. For the purpose of hierarchical analysis, countries and territories were organized into 21 regions, mostly based on geography and national income (Supplementary Table 1). Regions were in turn organized into nine super-regions.

We used observation year, that is, the year in which data were collected, as the timescale for the analysis of BMI and birth year as the timescale for the analysis of height, consistent with previous analyses[7,65,70]. Time trends were modelled through a combination of a linear term, to capture gradual long-term change, and a second-order random walk, which allows for nonlinear trends[71], both modelled hierarchically. The age associations of height and BMI were modelled, using cubic splines, to allow for nonlinear changes over age, including periods of rapid and slow rise. Periods of rapid rise representing adolescent growth spurts, which occur earlier in girls than boys[72,73,74], were reflected in the placement of spline knots for boys and girls, respectively, as detailed in the section on model specification. Spline coefficients were allowed to vary across countries, informed by their own data as well as data from other countries as specified by a hierarchical structure, as previously described[69].

The model also accounted for the possibility that height or BMI in subnational and community samples might differ systematically from nationally representative samples and have larger variation than in national studies. These features were accounted for through the inclusion of fixed-effect and random-effect terms for subnational and community data as detailed in the model specification section below. The fixed effects accounted for systematic differences between subnational or community studies and national studies. The inclusion of random effects allowed national data to have greater influence on the estimates than subnational or community data with similar sample sizes because the subnational and community data have additional variance from the random-effect terms. Both were estimated empirically as a part of model fitting.

Following the approach of previous papers[20,21,67], the model included parameters representing the urban–rural height or BMI difference, which is empirically estimated and allowed to vary by country and year. We further expanded the model to allow urban–rural difference in height or BMI to vary by age, as height or weight with age may vary between children residing in rural versus urban areas. If data for a country-year-age group contained a mix of children living in urban and rural areas but were not stratified by place of residence (21% of all data sources), the estimated height or BMI difference was informed by stratified data from other age groups, years and

countries, especially those in the same region with data from similar time periods and/or ages.

Statistical model specification

As stated earlier, for each data source, we calculated mean height and BMI, together with corresponding standard errors, stratified by sex, age and rural or urban place of residence. For sources that did not stratify the sample on the place of residence, we obtained age-and-sex-stratified data. Each study contributed up to 30 mean BMI data points or 32 mean height data points for each sex, with the exact number depending on how many age groups were represented in the study and whether the study provided data stratified on urban and rural place of residence. The likelihood for an observation at urbanicity level s (urban-only, rural-only or mixed; referred to as stratum hereafter) and age group h, with age z_h, from study i, carried out in country j at time t is as follows:

$$\begin{array}{c}{y}_{s,h,i}\sim N({a}_{j[i]}+{b}_{j[i]}{t}_{i}+{u}_{j[i],{t}_{i}}+{\gamma }_{i}({z}_{h})+{{\boldsymbol{X}}}_{{\boldsymbol{i}}}\,{\boldsymbol{\beta }}+{e}_{i}\\ \,\,+{I}_{s,i}[{p}_{j[i]}+{q}_{j[i]}{t}_{i}+{r}_{j[i]}{z}_{h}+{d}_{i}],{{\rm{S}}{\rm{D}}}_{s,h,i}^{2}/{n}_{s,h,i}+\,{\tau }^{2}),\end{array}$$

where the country-specific intercept and linear time slope from the jth country ($j = 1 \ldots J$, where $J = 200$ which is the total number of countries in our analysis) are denoted ${a}_{j}$ and ${b}_{j}$, respectively. We describe the hierarchical model used for the ${a}^{{\prime} }$s and ${b}^{{\prime} }s$ in the section ‘Linear components of country time trends’. Letting $T = 31$ be the total number of years from 1990 to 2020, the T-length vector ${u}_{j}$ captures smooth nonlinear change over time in country j, as described in the section ‘Nonlinear change’. The age effects of the hth age group (with age ${z}_{h}$) in study i are denoted by ${\gamma }_{i}$; we describe the age model in the section ‘Age model’. The matrix ${\boldsymbol{X}}$ contains terms describing whether studies were representative at the national, subnational or community level. In

addition, a random effect, e_{i}, is estimated for each study, described in the section ‘Study-level term and study-specific random effects’.

Linear components of country time trends

The model had a hierarchical structure, whereby studies were nested in countries, which were nested in regions (indexed by *k*), which were nested in super-regions (indexed by *l*), which were all nested in the globe (see Supplementary Table 1 for a list of countries and territories in each region, and regions in each super-region). This structure allowed the model to share information across units to a greater degree when data were non-existent or weakly informative (for example, had a small sample size or were not nationally representative) and, to a lesser extent, in data-rich countries and regions[75].

The a and b terms are country-specific linear intercepts and time slopes with terms at each level of the hierarchy, denoted by the superscripts *c*, *r*, *s* and *g*, respectively:

$$\begin{array}{l}\,{a}_{j}={a}_{j}^{c}+{a}_{k[\,j]}^{r}+{a}_{l[k]}^{s}+{a}^{{\rm{g}}},\\ \,{b}_{j}={b}_{j}^{c}+{b}_{k[\,j]}^{r}+{b}_{l[k]}^{s}+{b}^{{\rm{g}}},\\ {a}_{j}^{x}\sim N(0,{\kappa }_{a}^{x}),\\ {b}_{j}^{x}\sim N(0,{\kappa }_{b}^{x}),\end{array}$$

where $x = \{c, r, s\}$.

The κ terms were each assigned a flat prior on the s.d. scale[76]. We also assigned flat priors to ${a}^{{\rm{g}}}$ and ${b}^{{\rm{g}}}$.

Nonlinear change

Mean BMI or height may change nonlinearly over time[7,54,58,65,70]. We captured smooth nonlinear change in time in urban and rural strata of country *j* using the vector u_{j}. Just as a_{j} and b_{j} are each defined as the sum of country, region, super-region and global components, we defined

$${u}_{j}={u}_{j}^{c}+{u}_{k[\,j]}^{r}+{u}_{l[k]}^{s}+{u}^{{\rm{g}}}.$$

To allow the model to differentiate between the degrees of nonlinearity that exist at the country, region, super-region and global levels, we assigned the four components of each u a Gaussian autoregressive prior[71,77]. In particular, the T vectors ${u}_{j}^{c}$ ($j = 1 \ldots J$), ${u}_{k}^{r}$ ($k = 1 \ldots K$), ${u}_{l}^{s}$ ($l = 1 \ldots L$) and ${u}^{{\rm{g}}}$ each have a normal prior with mean zero and precision ${\lambda }_{{\rm{c}}}P$, ${\lambda }_{{\rm{r}}}P$, ${\lambda }_{{\rm{s}}}P$ and ${\lambda }_{{\rm{g}}}P$, respectively, where the scaled precision matrix P in the Gaussian autoregressive prior penalizes first and second differences as follows:

$$\begin{array}{c}P\,=\,[\begin{array}{ccccc}\,1 & \,0 & \,0 & \,\cdots & \,0\\ -2 & \,1 & \,0 & \,\cdots & \,0\\ \,1 & -2 & \,1 & \,\cdots & \,0\\ \,0 & \,1 & -2 & \,\cdots & \,0\\ \,0 & \,0 & \,1 & \,\cdots & \,0\\ \,\vdots & \,\vdots & \,\vdots & \,\ddots & \,\vdots \\ \,0 & \,0 & \,0 & \,\cdots & \,1\end{array}]\,[\begin{array}{ccccccc}\,1 & -2 & \,1 & \,0 & \,0 & \,\cdots & \,0\\ \,0 & \,1 & -2 & \,1 & \,0 & \,\cdots & \,0\\ \,0 & \,0 & \,1 & -2 & \,1 & \,\cdots & \,0\\ \,\vdots & \,\vdots & \,\vdots & \,\vdots & \,\vdots & \,\ddots & \,\vdots \\ \,0 & \,0 & \,0 & \,0 & \,0 & \,\cdots & \,1\end{array}]\\ \,=\,[\begin{array}{ccccccc}\,1 & -2 & \,1 & \,0 & \,0 & \,\cdots & \,0\\ -2 & \,5 & -4 & \,1 & \,0 & \,\cdots & \,0\\ \,1 & -4 & \,6 & -4 & \,1 & \,\cdots & \,0\\ \,0 & \,1 & -4 & \,6 & -4 & \,\cdots & \,0\\ \,0 & \,0 & \,1 & -4 & \,6 & \,\cdots & \,0\\ \,\vdots & \,\vdots & \,\vdots & \,\vdots & \,\vdots & \,\ddots & \,\vdots \\ \,0 & \,0 & \,0 & \,0 & \,0 & \,\cdots & \,1\end{array}].\end{array}$$

P is multiplied by the estimated precision parameters ${\lambda }_{{\rm{c}}}$, ${\lambda }_{{\rm{r}}}$, ${\lambda }_{{\rm{s}}}$ and ${\lambda }_{{\rm{g}}}$, thus upweighting or downweighting the strength of its penalties and ultimately determining the degree of smoothing at each level. For each of the four precision parameters, we used a truncated flat prior on the s.d. scale ($1/\surd \lambda$)[76]. We truncated these priors such that log$\lambda \le 20$ for each of the four ${\lambda }^{{\prime} }s$. This upper bound is enforced as a computational convenience, whereby models with log$\lambda > 20$ are treated as equivalent to a model with

$\log\lambda = 20$ as they essentially have no extralinear variability in time. In practice, this upper bound had little effect on the parameter estimates. Furthermore, we ordered the ${\lambda}^{\prime}s$ a priori as follows: ${\lambda}_{\rm{c}} < {\lambda}_{\rm{r}} < {\lambda}_{\rm{s}} < {\lambda}_{\rm{g}}$. This prior constraint conveys the natural expectation that, for example, the global height or BMI trend has less extralinear variability than the trend of any given region, which in turn has less variability than those of constituent countries.

The matrix P has rank $T - 2$, corresponding to a flat, improper prior on the mean and the slope of the ${u}_{j}^{\rm{c}}$'s, the ${u}_{k}^{\rm{r}}$'s and the ${u}_{l}^{\rm{s}}$'s and ${u}^{\rm{g}}$, and is not invertible[78]. Thus, we had a proper prior in a reduced-dimension space[71], with the prior expressed as follows:

$$P({u}_{j}^{\rm{c}}| {\lambda}_{\rm{c}})\propto {\lambda}_{\rm{c}}^{\frac{T-2}{2}}\exp \left\{-\frac{{\lambda}_{\rm{c}}}{2}{u}_{j}^{{\rm{c}}{\prime}}P{u}_{j}^{\rm{c}}\right\}.$$

Note that if ${u}_{j}^{\rm{c}}$ had a non-zero mean, this would introduce non-identifiability with respect to ${a}_{j}^{\rm{c}}$. By the same token, ${b}_{j}^{\rm{c}}$ would not be identifiable if ${u}_{j}$ had a non-zero time slope, and similarly for the other means and slopes. Thus, to achieve identifiability of the ${a}^{\prime}s$, ${b}^{\prime}s$, and ${u}^{\prime}{\rm{s}}$, we constrained the mean and slope of ${u}^{\rm{g}}$ and of each ${u}^{\rm{s}}$, ${u}^{\rm{r}}$ and ${u}^{\rm{c}}$ to be zero. Enforcing orthogonality between the linear and nonlinear portions of the time trends meant that each can be interpreted independently.

For the cases in which we have observations for at least two different time points, this improper prior will not lead to an improper posterior because the data will provide information about the mean and slope. In order to enforce the desired orthogonality between the linear and nonlinear portions of the model, we constrained the mean and slope of the ${u}_{j}^{\rm{c}}$'s, ${u}_{k}^{\rm{r}}$'s and ${u}_{l}^{\rm{s}}$'s and of ${u}^{\rm{g}}$ to be zero[71].

For the six countries with no height data, and seven countries with no BMI data, we took the Moore–Penrose pseudoinverse of *P* [79], setting to infinity those eigenvalues that correspond to the non-identifiability. This effectively constrained the non-identified portions of the model to zero, as the corresponding variances are set to zero[77]; in this case the Rue and Held correction[71] is not needed. An intermediate case occurs when data are observed for only one time point in a country. In this case, the full conditional precision has rank $T-1$ because the mean but not the linear trend of ${u}_{j}^{{\rm{c}}}$ is identified by the data. We therefore constrained the linear trend of ${u}_{j}^{{\rm{c}}}$ to zero by taking the generalized inverse of the full conditional precision. We then constrained the mean of ${u}_{j}^{{\rm{c}}}$ to zero using the one-dimensional version of the Rue and Held correction[71].

Age model

To capture sex-specific patterns of growth, especially adolescent growth spurts, we modelled age using cubic splines. The number and position of the knots of the splines were selected on the basis of a combination of physiological and statistical considerations, as described in a national level analysis[7]. For age group *h* with age z_h, in study *i*, the age effect for height and BMI is given, respectively, as follows:

$$\begin{array}{l}{\gamma }_{i}({z}_{h})={\gamma }_{1i}\,{z}_{h}+{\gamma }_{2i}{z}_{h}^{2}+{\gamma }_{3i}{z}_{h}^{3}+{\gamma }_{4i}{({z}_{h}-{k}_{1})}_{+}^{3}+{\gamma }_{5i}{({z}_{h}-{k}_{2})}_{+}^{3}+{\gamma }_{6i}{({z}_{h}-{k}_{3})}_{+}^{3}+{\gamma }_{7i}{({z}_{h}-{k}_{4})}_{+}^{3},\end{array}$$

(height)

$$\begin{array}{l}{\gamma }_{i}({z}_{h})={\gamma }_{1i}\,{z}_{h}+{\gamma }_{2i}{z}_{h}^{2}+{\gamma }_{3i}{z}_{h}^{3}+{\gamma }_{4i}{({z}_{h}-{k}_{1})}_{+}^{3}+{\gamma }_{5i}{({z}_{h}-{k}_{2})}_{+}^{3}.\end{array}$$

(BMI)

For height, four spline knots were placed at ages $\{{k}_{1},{k}_{2},{k}_{3},{k}_{4}\}\,=\{\,8,\,10,\,12,14\}$ for girls and at ages {\

${k}_{1}$, ${k}_{2},{k}_{3},{k}_{4}=\{10,12,14,16\}$ for boys. For BMI, we used two spline knots (at ages 10 and 15 years) because, at the population level, changes in BMI with age are smoother than those in height[7,72,73]. Each of the spline coefficients was allowed to vary across countries, with a hierarchical structure as described in a previous paper[69], using the equation below, where ψ is the global intercept, and $c,r\,{\rm{and}}\,s$ are the country, region and super-region random intercepts, respectively. The *k*th age effect coefficients for study *i* (${\gamma }_{k,i}$) for each age group *h*, with age z_h, are given as follows:

$$\begin{array}{l}{\gamma }_{k,i}={\psi }_{k}+{c}_{k,j[i]}+{r}_{k,l[i]}+{s}_{k,m[i]},\\ {c}_{k,j} \sim N(0,{\sigma }_{k,c}^{2}),\\ {r}_{k,l} \sim N(0,{\sigma }_{k,r}^{2}),\\ {s}_{k,l} \sim N(0,{\sigma }_{k,s}^{2}).\end{array}$$

A flat improper prior was placed on each of the ${{\sigma }}_{k}$'s .

Study-level term and study-specific random effects

Mean height or BMI from individual studies may deviate from the true country-year mean owing to factors associated with sampling, response or measurement. We used a study-level term to help account for potential systematic differences associated with data sources that are representative of subnational and community populations. Our model therefore included time-varying offsets (referred to as fixed effects above) for subnational and community data in the term ${{\boldsymbol{X}}}_{{\boldsymbol{i}}}{\boldsymbol{\beta }}$:

$$\begin{array}{l}{{\boldsymbol{X}}}_{{\boldsymbol{i}}}{\boldsymbol{\beta }}={\beta }_{1}I\{{X}_{j[i],t[i]}^{{\rm{cvrg}}}={\rm{subnational}}\}+{\beta }_{2}I\{{X}_{j[i],t[i]}^{{\rm{cvrg}}}={\rm{subnational}}\}{t}_{i}\\ \,+\,{\beta }_{3}I\{{X}_{j[i],t[i]}^{{\rm{cvrg}}}={\rm{community}}\}+{\beta }_{4}I\{{X}_{j[i],t[i]}^{{\rm{cvrg}}}={\rm{community}}\}{t}_{i},\end{array}$$

where ${X}_{j\left[i\right],t\left[i\right]}^{{\rm{cvrg}}}$ is the indicator for whether the coverage of study *i*, in country *j* and year *t*, is subnational or community.

Even after accounting for sampling variability, national studies may still not reflect the true mean height or BMI level of a country with perfect accuracy, and subnational and community studies have even larger variability. In study *i*, the study-specific random effect ${e}_{i}$ allows all age groups from the same study to have an unusually high or an unusually low mean after conditioning on the other terms in the model. Each ${e}_{i}$ is assigned a normal prior with variance depending on whether study *i* is representative at the national, subnational or community level. Random effects from national studies were constrained to have smaller variance (${v}_{{\rm{n}}}$) than random effects of subnational studies (${v}_{{\rm{s}}}$), which were in turn constrained to have smaller variance than community studies (${v}_{{\rm{c}}}$). To make country-level predictions, we set ${e}_{i}=0$, thus not including random effects arising from imperfections and variations in study design and implementation and from within-country variability of height or BMI means.

Urban and rural strata

To model mean height and BMI by urban and rural places of residence, the model included offsets for the two strata. The offsets were captured by country-specific intercept, linear time and age effects, using a centred indicator term (${I}_{s,i}$):

$${I}_{s,i}[{p}_{j[i]}+{q}_{j[i]}{t}_{i}+{r}_{j[i]}{z}_{h}+{d}_{i}],$$

where ${I}_{s,i}=-1+2{X}_{s,i}^{{\rm{u}}{\rm{r}}{\rm{b}}}$, with

$${X}_{s,i}^{{\rm{urb}}}=\left\{\begin{array}{ll}1, & {\rm{if}}\,{\rm{stratum}}\,s\,{\rm{contains}}\,{\rm{only}}\,{\rm{urban}}\,{\rm{individuals}},\\ 0, & {\rm{if}}\,{\rm{stratum}}\,s\,{\rm{contains}}\,{\rm{only}}\,{\rm{rural}}\,{\rm{individuals}},\\ {X}_{j[i],t[i]}^{{\rm{urb}}}, & {\rm{if\; stratum}}\,s\,{\rm{contains\; a\; mixture\; of\; urban}}\,{\rm{and}}\,{\rm{rural}}\,{\rm{individuals}}.\end{array}\right.$$

In other words, for data not stratified by place of residence, the model treated the unstratified mean height or BMI as equivalent to the weighted sum of the (unobserved) urban sample mean height or BMI and rural sample mean height or BMI, with the weights based on the proportion of the population of that country living in urban areas in the year of the survey (${X}_{j\left[i\right],t[i]}^{{\rm{urb}}}$).

The intercept (p) and slope (q) terms capture the country-to-country variation in the magnitude of the height or BMI difference between urban and rural populations and how the difference changes over time. The slope (r) captures the country-to-country variation in the BMI or height difference between urban and rural populations across age groups. These were specified with the same geographical hierarchy as the country-specific intercepts (a) and slopes (b) as follows:

$$\begin{array}{l}\,{p}_{j}={p}_{j}^{c}+{p}_{k[\,j]}^{r}+{p}_{l[k]}^{s}+{p}^{{\rm{g}}},\\ \,{q}_{j}={q}_{j}^{c}+{q}_{k[\,j]}^{r}+{q}_{l[k]}^{s}+{q}^{{\rm{g}}},\\ \,{{\rm{r}}}_{j}={r}_{j}^{c}+{r}_{k[\,j]}^{r}+{r}_{l[k]}^{s}+{r}^{{\rm{g}}},\\ {p}_{j}^{x}\sim N(0,{\kappa }_{p}^{x}),\\ {q}_{j}^{x}\sim N(0,{\kappa }_{q}^{x}),\\ {r}_{j}^{x}\sim N(0,{\kappa }_{r}^{x}),\end{array}$$

where $x=\{c,r,s\}$. The study random effect term ${d}_{i}$ incorporates deviations from the country-level urban–rural difference in each study and is analogous to ${e}_{i}$.

Residual age-by-study variability

The age patterns across communities within a given country may differ from the overall age pattern of that country. This within-study variability cannot be captured by the e terms, which are equal across age-specific observations in each study, so we included an additional variance component for each study, ${\tau }^{2}$.

Model implementation

All analyses were done separately by sex because age, geographical and temporal patterns of height and BMI differ between girls and boys[7,65]. We fitted the statistical model using Markov chain Monte Carlo (MCMC). We started 35 parallel MCMC runs from randomly generated overdispersed starting values. For computational efficiency, each chain was run for a total of 75,000 iterations. All chains converged to the same target distribution within this number, but due to the overdispersed initial values, the length of burn-in required to converge to the target distribution varied. After the runs were completed, we used trace plots to monitor convergence and to select chains that had completed burn-in within 35,000 iterations. This resulted in 16 chains for boys and 17 for girls for BMI, and 14 chains for boys and 16 for girls for height. Within each of these chains, post-burn-in iterations were thinned by keeping every 10th iteration, which were then combined for all chains and further thinned to a final set of 5,000 draws of the model parameter estimates. We used the posterior distribution of the model parameters to obtain the posterior distributions of our outcomes: mean urban and rural height and BMI, and the urban–rural difference in mean height and BMI. Posterior estimates were made for one-year age groups from 5 to 19 years, as well as for age-standardized outcomes, by year. The reported CrIs represent the 2.5th and the 97.5th percentiles of the posterior distributions. We also report the posterior s.d. of estimates, and PP that the estimated change in height or BMI in rural or urban areas, and in the urban–rural height or BMI difference over time, represents a true increase or decrease.

Convergence was confirmed for the country-sex specific posterior outcomes —namely mean urban height and BMI, mean rural height and BMI and the urban–rural difference in mean height and BMI—for reporting ages (5, 10, 15, 19 years and age-standardized) and years (1990 and 2020) using the R-hat diagnostic[80,81]. For height, the 2.5th to 97.5th percentiles of the R-hats for the reporting ages and years were 0.999–1.010 for girls and 0.999–1.004 for boys. For BMI, the 2.5th to 97.5th percentiles of the R-hats were 0.999–1.004 for girls and 0.999–1.005 for boys.

We applied the pool-adjacent-violators algorithm, a monotonic regression that uses an iterative algorithm based on least squares to fit a free-form line to a sequence of observations such that the fitted line is non-decreasing[82,83], on the posterior height estimates to ensure that the height for each birth

cohort increased monotonically with age. In practice, this had little effect on the results, with height at age 19 years adjusted by an average of 0.26 cm or less for both boys and girls. All analyses were conducting using the statistical software R (v.4.1.2)[84].

Strengths and limitations

An important strength of our study is its novel scope of presenting consistent and comparable estimates of urban and rural height and BMI among school-aged children and adolescents, which is essential to formulate and evaluate policies that aim to improve health in these formative ages. We used a large number of population-based studies from 194 countries and territories covering around 99% of the population of the world. We maintained a high level of data quality and representativeness through repeated checks of study characteristics against our inclusion and exclusion criteria, and did not use any self-reported data to avoid bias in height and weight. Data were analysed according to a consistent protocol, and the characteristics and quality of data from each country were rigorously verified through repeated checks by NCD-RisC members. We used a statistical model that used all available data and took into account the epidemiological features of height and BMI during childhood and adolescence by using nonlinear time trends and age associations. The model used the available information on the urban–rural difference in height and BMI and estimated the age-varying and time-varying urban–rural difference for all countries and territories hierarchically.

Despite our extensive efforts to identify and access data, some countries had fewer data, especially those in the Caribbean, Polynesia, Micronesia and sub-Saharan Africa. Of the studies used, fewer than half had data for children aged 5–9 years compared to nearly 90% with data for children and adolescents aged 10–19 years. The scarcity of data is reflected in the larger uncertainty of our estimates for these countries and regions, and younger age groups. This reflects the need to systematically include school-aged children in both health and nutrition surveys, and, especially in countries where school enrolment is high, to use schools as a platform for monitoring growth and developmental outcomes for entire national populations and key subgroups such as those in rural and urban areas. Although urban and rural

classifications are commonly based on definitions by national statistical offices, classification of cities and rural areas may, appropriately, vary by country according to their demographic characteristics (for example, population size or density), economic activities, administrative structures, infrastructure and environment. Similarly, urbanization takes place through a variety of mechanisms such as changes in fertility in rural and urban areas, migration and reclassification of previously rural areas to urban as they grow and industrialize. Each of these mechanisms may have different implications for nutrition and physical activity, and hence height and/or BMI, and should be a subject of studies that follow individual participants and changes in their place of residence. Finally, there is variation in growth and development of children within rural or urban areas based on household socioeconomic status and community characteristics that affect access to and the quality of nutrition, the living environment and healthcare[35,85,86]. Among these, in some cities, a large number of families live in slums[19,87]. School-aged children and adolescents living in slums have nutrition, environment and healthcare access that is typically worse than other residents of the city, although often better than those in rural areas[19,87,88,89,90].

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Estimates of mean BMI and height by country, year, sex, single year of age as well as age-standardized, and place of residence (urban and rural) will be available from https://www.ncdrisc.org in machine-readable numerical format and as visualizations upon publication of the paper. Input data from publicly available sources and contact information for data providers can be downloaded from https://www.ncdrisc.org and Zenodo (https://doi.org/10.5281/zenodo.7355601).

Code availability

The computer code for the Bayesian hierarchical model and the code used to generate figures in this work will be available at https://www.ncdrisc.org and Zenodo (https://doi.org/10.5281/zenodo.7355601).

References

1. Prentice, A. M. et al. Critical windows for nutritional interventions against stunting. *Am. J. Clin. Nutr.* **97**, 911–918 (2013).

2. Tanner, J. M. Growth as a mirror of the condition of society: secular trends and class distinctions. *Acta Paediatr. Jpn.* **29**, 96–103 (1987).

3. Strauss, J. & Thomas, D. Health, nutrition, and economic development. *J. Econ. Lit.* **36**, 766–817 (1998).

4. Nüesch, E. et al. Adult height, coronary heart disease and stroke: a multi-locus Mendelian randomization meta-analysis. *Int. J. Epidemiol.* **45**, 1927–1937 (2016).

5. Park, M. H., Falconer, C., Viner, R. M. & Kinra, S. The impact of childhood obesity on morbidity and mortality in adulthood: a systematic review. *Obes. Rev.* **13**, 985–1000 (2012).

6. Caird, J. et al. Does being overweight impede academic attainment? A systematic review. *Health Educ. J.* **73**, 497–521 (2014).

7. NCD Risk Factor Collaboration (NCD-RisC). Height and body-mass index trajectories of school-aged children and adolescents from 1985 to 2019 in 200 countries and territories: a pooled analysis of 2181 population-based studies with 65 million participants. *Lancet* **396**, 1511–1524 (2020).

8. *Global Report on Urban Health: Equitable Healthier Cities for Sustainable Development* (World Health Organization & UN-Habitat, 2016).

9. *Healthy Environments for Healthy Children: Key Messages for Action. Report No. 924159988X* (World Health Organization, 2010).

10. *Report of the Commission on Ending Childhood Obesity* (World Health Organization, 2016).

11. Smith, D. M. & Cummins, S. Obese cities: how our environment shapes overweight. *Geogr. Compass* **3**, 518–535 (2009).

12. Fraser, B. Latin America's urbanisation is boosting obesity. *Lancet* **365**, 1995–1996 (2005).

13. Pirgon, Ö. & Aslan, N. The role of urbanization in childhood obesity. *J. Clin. Res. Pediatr. Endocrinol.* **7**, 163–167 (2015).

14. *The State of the World's Children 2012: Children in an Urban World* (UNICEF, 2012).

15. Kirchengast, S. & Hagmann, D. "Obesity in the city"—urbanization, health risks and rising obesity rates from the viewpoint of human biology and public health. *Hum. Biol. Public Health.* **2**, 1–21 (2021).

16. Congdon, P. Obesity and urban environments. *Int. J. Environ. Res.* **16**, 464 (2019).

17. Gong, P. et al. Urbanisation and health in China. *Lancet* **379**, 843–852 (2012).

18. Liese, A. D., Weis, K. E., Pluto, D., Smith, E. & Lawson, A. Food store types, availability, and cost of foods in a rural environment. *J. Am. Diet. Assoc.* **107**, 1916–1923 (2007).

19. Krumdiek, C. L. The rural-to-urban malnutrition gradient: a key factor in the pathogenesis of urban slums. *JAMA* **215**, 1652–1654 (1971).

20. Paciorek, C. J., Stevens, G. A., Finucane, M. M. & Ezzati, M. Children's height and weight in rural and urban populations in low-income and middle-income countries: a systematic analysis of population-representative data. *Lancet Glob. Health* **1**, e300–e309 (2013).

21. NCD Risk Factor Collaboration (NCD-RisC). Rising rural body-mass index is the main driver of the global obesity epidemic in adults. *Nature* **569**, 260–264 (2019).

22. Jaacks, L. M., Slining, M. M. & Popkin, B. M. Recent trends in the prevalence of under- and overweight among adolescent girls in low- and middle-income countries. *Pediatr. Obes.* **10**, 428–435 (2015).

23. Wang, Y., Monteiro, C. & Popkin, B. M. Trends of obesity and underweight in older children and adolescents in the United States, Brazil, China, and Russia. *Am. J. Clin. Nutr.* **75**, 971–977 (2002).

24. Neuman, M., Kawachi, I., Gortmaker, S. & Subramanian, S. V. Urban–rural differences in BMI in low- and middle-income countries: the role of socioeconomic status. *Am. J. Clin. Nutr.* **97**, 428–436 (2013).

25. *Hidden Cities: Unmasking and Overcoming Health Inequities in Urban Settings. Report No. 9241548037* (World Health Organization & Centre for Health Development, 2010).

26. *Population-based Approaches to Childhood Obesity Prevention. Report No. 9789241504782* (World Health Organization, 2012).

27. *Prevention of Overweight and Obesity in Children and Adolescents: UNICEF Programming Guidance* (UNICEF, 2019).

28. de Sa, T. H. et al. Urban design is key to healthy environments for all. *Lancet Glob. Health* **10**, e786–e787 (2022).

29. Kelly, C., Callaghan, M., Molcho, M., Nic Gabhainn, S. & Alforque Thomas, A. Food environments in and around post-primary schools in Ireland: associations with youth dietary habits. *Appetite* **132**, 182–189 (2019).

30. Conrad, D. & Capewell, S. Associations between deprivation and rates of childhood overweight and obesity in England, 2007–2010: an ecological study. *BMJ Open* **2**, e000463 (2012).

31. Katsoulis, M. et al. Identifying adults at high-risk for change in weight and BMI in England: a longitudinal, large-scale, population-based cohort study using electronic health records. *Lancet Diabetes Endocrinol.* **9**, 681–694 (2021).

32. Arnett, J. J. Emerging adulthood: a theory of development from the late teens through the twenties. *Am. Psychol.* **55**, 469–480 (2000).

33. Poobalan, A. S., Aucott, L. S., Clarke, A. & Smith, W. C. S. Diet behaviour among young people in transition to adulthood (18–25 year olds): a mixed method study. *Health Psychol. Behav. Med.* **2**, 909–928 (2014).

34. Winpenny, E. M. et al. Changes in diet through adolescence and early adulthood: longitudinal trajectories and association with key life transitions. *Int. J. Behav. Nutr. Phys. Act.* **15**, 86 (2018).

35. Smith, L. C., Ruel, M. T. & Ndiaye, A. Why is child malnutrition lower in urban than in rural areas? Evidence from 36 developing countries. *World Develop.* **33**, 1285–1305 (2005).

36. Lundborg, P., Rooth, D.-O. & Alex-Petersen, J. Long-term effects of childhood nutrition: evidence from a school lunch reform. *Rev. Econ. Stud.* **89**, 876–908 (2022).

37. Graça, P., Gregório, M. J. & Freitas, M. G. A decade of food and nutrition policy in Portugal (2010–2020). *Port. J. Public Health* **38**, 94–118 (2020).

38. Winichagoon, P. Thailand nutrition in transition: situation and challenges of maternal and child nutrition. *Asia Pac. J. Clin. Nutr.* **22**, 6–15 (2013).

39. Paes-Sousa, R. & Vaitsman, J. The Zero Hunger and Brazil without Extreme Poverty programs: a step forward in Brazilian social protection policy. *Cien Saude Colet*. **19**, 4351–4360 (2014).

40. Núñez, J. & Pérez, G. The escape from malnutrition of Chilean boys and girls: height-for-age *Z* scores in late XIX and XX centuries. *Int. J.*

Environ. Res. Public Health **18**, 10436 (2021).

41. Herrmann, R., Möser, A. & Weber, S. A. Grocery retailing in Germany: Situation, development and pricing strategies PDF Logo (Justus-Liebig-Universität Gießen, 2009).

42. Thang, N. M. & Popkin, B. Child malnutrition in Vietnam and its transition in an era of economic growth. *J. Hum. Nutr. Diet.* **16**, 233–244 (2003).

43. Subramanyam, M. A., Kawachi, I., Berkman, L. F. & Subramanian, S. V. Socioeconomic inequalities in childhood undernutrition in India: analyzing trends between 1992 and 2005. *PLoS ONE* **5**, e11392 (2010).

44. Jingzhong, Y. & Lu, P. Differentiated childhoods: impacts of rural labor migration on left-behind children in China. *J. Peasant Stud.* **38**, 355–377 (2011).

45. Sundberg, S. Agriculture, poverty and growth in Africa: linkages and policy challenges. *CABI Rev.* (2009).

46. Bentham, J. et al. Multidimensional characterization of global food supply from 1961 to 2013. *Nat. Food* **1**, 70–75 (2020).

47. Micha, R. et al. Global, regional and national consumption of major food groups in 1990 and 2010: a systematic analysis including 266 country-specific nutrition surveys worldwide. *BMJ Open* **5**, e008705 (2015).

48. Fraval, S. et al. Food access deficiencies in sub-Saharan Africa: prevalence and implications for agricultural interventions. *Front. Sustain. Food Syst.* **3**, 104 (2019).

49. Thomson, M., Kentikelenis, A. & Stubbs, T. Structural adjustment programmes adversely affect vulnerable populations: a systematic-narrative review of their effect on child and maternal health. *Public Health Rev.* **38**, 13 (2017).

50. Frisch, R. E. & Revelle, R. Height and weight at menarche and a hypothesis of critical body weights and adolescent events. *Science* **169**, 397–399 (1970).

51. Holmgren, A. et al. Pubertal height gain is inversely related to peak BMI in childhood. *Pediatr. Res.* **81**, 448–454 (2017).

52. Alhassan, J. A. K., Adeyinka, D. A. & Olakunde, B. O. Equity dimensions of the decline in under-five mortality in Ghana: a joinpoint regression analysis. *Tropi. Med. Int. Health* **25**, 732–739 (2020).

53. Wolde, K. S. & Bacha, R. H. Trend and correlates of under-5 mortality in Ethiopia: a multilevel model comparison of 2000–2016 EDHS data. *SAGE Open Med.* **10**, 20503121221100608 (2022).

54. NCD Risk Factor Collaboration (NCD-RisC). Trends in adult body–mass index in 200 countries from 1975 to 2014: a pooled analysis of 1698 population-based measurement studies with 19·2 million participants. *Lancet* **387**, 1377–1396 (2016).

55. Danaei, G. et al. National, regional, and global trends in systolic blood pressure since 1980: systematic analysis of health examination surveys and epidemiological studies with 786 country-years and 5·4 million participants. *Lancet* **377**, 568–577 (2011).

56. Danaei, G. et al. National, regional, and global trends in fasting plasma glucose and diabetes prevalence since 1980: systematic analysis of health examination surveys and epidemiological studies with 370 country-years and 2·7 million participants. *Lancet* **378**, 31–40 (2011).

57. Farzadfar, F. et al. National, regional, and global trends in serum total cholesterol since 1980: systematic analysis of health examination surveys and epidemiological studies with 321 country-years and 3·0 million participants. *Lancet* **377**, 578–586 (2011).

58. Finucane, M. M. et al. National, regional, and global trends in body-mass index since 1980: systematic analysis of health examination

surveys and epidemiological studies with 960 country-years and 9·1 million participants. *Lancet* **377**, 557–567 (2011).

59. Cline, M. G., Meredith, K. E., Boyer, J. T. & Burrows, B. Decline of height with age in adults in a general population sample: estimating maximum height and distinguishing birth cohort effects from actual loss of stature with aging. *Hum. Biol.* **61**, 415–425 (1989).

60. Lumley, T. Analysis of complex survey samples. *J. Stat. Softw.* **9**, 1–19 (2004).

61. United Nations Department of Economic and Social Affairs, Population Division. *World Urbanization Prospects: The 2018 Revision* (United Nations, 2019).

62. Hayes, A. J., Clarke, P. M. & Lung, T. W. Change in bias in self-reported body mass index in Australia between 1995 and 2008 and the evaluation of correction equations. *Popul. Health Metr.* **9**, 53 (2011).

63. Connor Gorber, S., Tremblay, M., Moher, D. & Gorber, B. A comparison of direct vs. self-report measures for assessing height, weight and body mass index: a systematic review. *Obes. Rev.* **8**, 307–326 (2007).

64. Ezzati, M., Martin, H., Skjold, S., Vander Hoorn, S. & Murray, C. J. L. Trends in national and state-level obesity in the USA after correction for self-report bias: analysis of health surveys. *J. R. Soc. Med.* **99**, 250–257 (2006).

65. NCD Risk Factor Collaboration (NCD-RisC). Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128·9 million children, adolescents, and adults. *Lancet* **390**, 2627–2642 (2017).

66. Ahmad, O. B. et al. *Age Standardization of Rates: A New WHO Standard* (World Health Organization, 2001).

67. Finucane, M. M., Paciorek, C. J., Stevens, G. A. & Ezzati, M. Semiparametric Bayesian density estimation with disparate data sources: a meta-analysis of global childhood undernutrition. *J. Am. Stat. Assoc.* **110**, 889–901 (2015).

68. Finucane, M. M., Paciorek, C. J., Danaei, G. & Ezzati, M. Bayesian estimation of population-level trends in measures of health status. *Stat. Sci.* **29**, 18–25 (2014).

69. NCD Risk Factor Collaboration (NCD-RisC). Worldwide trends in hypertension prevalence and progress in treatment and control from 1990 to 2019: a pooled analysis of 1201 population-representative studies with 104 million participants. *Lancet* **398**, 957–980 (2021).

70. NCD Risk Factor Collaboration (NCD-RisC). A century of trends in adult human height. *eLife* **5**, e13410 (2016).

71. Rue, H. V. & Held, L. *Gaussian Markov Random Fields: Theory and Applications* (Chapman & Hall/CRC, 2005).

72. Cameron, N. L. *Human Growth and Development* (Academic Press, 2002).

73. de Onis, M. et al. Development of a WHO growth reference for school-aged children and adolescents. *Bull. World Health Organ.* **85**, 660–667 (2007).

74. Tanner, J. M., Whitehouse, R. H., Marubini, E. & Resele, L. F. The adolescent growth spurt of boys and girls of the Harpenden growth study. *Ann. Hum. Biol.* **3**, 109–126 (1976).

75. Gelman, A. & Jill, J. *Data Analysis Using Regression and Multilevel/Hierarchical Models* (Cambridge Univ. Press, 2007).

76. Gelman, A. Prior distributions for variance parameters in hierarchical models. *Bayesian Anal.* **1**, 515–533 (2006).

77. Breslow, N. E. & Clayton, D. G. Approximate inference in generalized linear mixed models. *J. Am. Stat. Assoc.* **88**, 9–25 (1993).

78. Wood, S. N. *Generalized Additive Models: An Introduction with R* (Chapman & Hall/CRC, 2006).

79. Harville, D. A. *Matrix Algebra from a Statistician's Perspective* (Springer-Verlag, 2008).

80. Vehtari, A., Gelman, A., Simpson, D., Carpenter, B. & Bürkner, P.-C. Rank-normalization, folding, and localization: an improved R for assessing convergence of MCMC (with discussion). *Bayesian Anal.* **16**, 667–718 (2021).

81. Stan Development Team. *RStan: the R interface to Stan* https://mc-stan.org/ (2022).

82. Best, M. J. & Chakravarti, N. Active set algorithms for isotonic regression; a unifying framework. *Math. Program.* **47**, 425–439 (1990).

83. De Leeuw, J., Hornik, K. & Mair, P. Isotone optimization in R: pool-adjacent-violators algorithm (PAVA) and active set methods. *J. Stat. Softw.* **32**, 1–24 (2010).

84. The R Development Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2021).

85. Huang, R., Moudon, A. V., Cook, A. J. & Drewnowski, A. The spatial clustering of obesity: does the built environment matter? *J. Hum. Nutr. Diet.* **28**, 604–612 (2015).

86. Menon, P., Ruel, M. T. & Morris, S. S. Socio-economic differentials in child stunting are consistently larger in urban than in rural areas. *Food Nutr. Bull.* **21**, 282–289 (2000).

87. Ezeh, A. et al. The history, geography, and sociology of slums and the health problems of people who live in slums. *Lancet* **389**, 547–558 (2017).

88. Fink, G., Günther, I. & Hill, K. Slum residence and child health in developing countries. *Demography* **51**, 1175–1197 (2014).

89. Jonah, C. M. P. & May, J. D. The nexus between urbanization and food insecurity in South Africa: does the type of dwelling matter? *Int. J. Urban Sustain. Dev.* **12**, 1–13 (2020).

90. *Progress Report: Sustainable Development Goal 11 (Target 11.1)* (Habitat for Humanity International, 2019).

Acknowledgements

This study was funded by the UK Medical Research Council (grant number MR/V034057/1), the Wellcome Trust (Pathways to Equitable Healthy Cities grant 209376/Z/17/Z), the AstraZeneca Young Health Programme and the European Commission (STOP project through EU Horizon 2020 research and innovation programme under Grant Agreement 774548). For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) licence to the Author Accepted Manuscript version arising from this submission. We thank W. Dietz, L. Jaacks and W. Johnson for recommendations of relevant citations. The authors alone are responsible for the views expressed in this Article and they do not necessarily represent the views, decisions, or policies of the institutions with which they are affiliated.

Author information

Author notes

1. These authors contributed equally: Anu Mishra, Bin Zhou, Andrea Rodriguez-Martinez, Honor Bixby

2. Deceased: Konrad Jamrozik, Altan Onat, Robespierre Ribeiro, Michael Sjöström, Agustinus Soemantri, Jutta Stieber, Dimitrios Trichopoulos

Authors and Affiliations

1. Imperial College London, London, UK

   Anu Mishra, Bin Zhou, Andrea Rodriguez-Martinez, Rosie K. Singleton, Rodrigo M. Carrillo-Larco, Kate E. Sheffer, James E. Bennett, Victor Lhoste, Maria L. C. Iurilli, Nowell H. Phelps, Marisa K. Sophiea, Edward W. Gregg, Queenie Chan, Amanda J. Cross, Paul

Elliott, Ilpo Tapani Huhtaniemi, Marjo-Riitta Jarvelin, Elio Riboli, Evangelia Tzala, Mark Woodward & Majid Ezzati

2. McGill University, Montreal, Québec, Canada

Honor Bixby, David Goltzman & Suzanne N. Morin

3. University of Essex, Colchester, UK

Honor Bixby & Mariachiara Di Cesare

4. University of California Berkeley, Berkeley, CA, USA

Christopher J. Paciorek & Farnam Mohebi

5. University of Kent, Canterbury, UK

James Bentham

6. World Health Organization, Geneva, Switzerland

Gretchen A. Stevens, Melanie J. Cowan, Stefan Savin, Leanne M. Riley, Mohamed M. Ali & Francesco Branca

7. Harvard T. H. Chan School of Public Health, Boston, MA, USA

Goodarz Danaei, Wafaie W. Fawzi, Yanping Li, Caleb Ochimana & Dimitrios Trichopoulos

8. Mahidol University, Nakhon Pathom, Thailand

Wichai Aekplakorn, Nipa Rojroongwasinkul, Paibul Suriyawongpaisal & Uruwan Yamborisut

9. Ministry of Health, Kuala Lumpur, Malaysia

Noor Ani Ahmad, Suhaila Abdul Ghaffar, Mohamad Hasnan Ahmad, Tahir Aris, Azli Baharudin, Norsyamlina Che Abdul Rahim, Norazizah Ibrahim Wong, Muhammad Fadhli Mohd

Yusoff, Mohd Azahadi Omar, Rusidah Selamat, Ahmad Faudzi Yusoff & Ahmad A. Zainuddin

10. Bispebjerg and Frederiksberg Hospital, Copenhagen, Denmark

Jennifer L. Baker, Kristine Allin, Anne A. Bjerregaard, Thomas M. Dantoft, Marie Eliasen, Torben Jørgensen, Freja B. Kampmann, Line L. Kårhus, Katja B. Leth-Møller, Allan Linneberg, Anja L. Madsen, Line T. Møllehave, Merete Osler & Betina H. Thuesen

11. Victor Babes University of Medicine and Pharmacy, Timisoara, Romania

Adela Chirita-Emandi, Mihai Gafencu, Iulia Jurca Simina & Maria Puiu

12. Non-Communicable Diseases Research Center, Tehran, Iran

Farshad Farzadfar, Naser Ahmadi, Mehrdad Azmin, Ali Ghanbari, Erfan Ghasemi, Rosa Haghshenas, Parinaz Mehdipour, Sahar Saeedi Moghaddam, Bahram Mohajer, Shohreh Naderimagham & Moein Yoosefi

13. Swiss Tropical and Public Health Institute, Basel, Switzerland

Günther Fink

14. University of Basel, Basel, Switzerland

Günther Fink

15. World Health Organization Regional Office for Europe, Copenhagen, Denmark

Mirjam Heinen, Marta Buoncristiano, Ivo Rakovac, Kremlin Wickramasinghe & Julianne Williams

16. National Institutes of Biomedical Innovation, Health and Nutrition, Tokyo, Japan

Nayu Ikeda & Nobuo Nishi

17. South African Medical Research Council, Cape Town, South Africa

Andre P. Kengne

18. Seoul National University College of Medicine, Seoul, Republic of Korea

Young-Ho Khang

19. University of Eastern Finland, Kuopio, Finland

Tiina Laatikainen, Sari Hantunen, Jussi Kauhanen, Jaakko Mursu, Tomi-Pekka Tuomainen, Jyrki K. Virtanen & Ari Voutilainen

20. Finnish Institute for Health and Welfare, Helsinki, Finland

Tiina Laatikainen, Tuija Jääskeläinen, Anne Juolevi, Eero O. Kajantie, Seppo Koskinen, Kari Kuulasmaa, Annamari Lundqvist, Päivi Mäki, Markku Peltonen, Veikko Salomaa, Hanna K. Tolonen & Jaakko Tuomilehto

21. ICMR–National Institute of Nutrition, Hyderabad, India

Avula Laxmaiah, Nimmathota Arlappa, Nagalla Balakrishna, Rachakulla Hari Kumar, Kodavanti Mallikharjuna Rao & Indrapal I. Meshram

22. Peking University, Beijing, China

Jun Ma, Yanhui Dong, Pei Gao, Yonghua Hu, Guansheng Ma, Yi Song, Xun Tang, Yong Tao & Yi Zeng

23. Universidad de San Carlos, Guatemala City, Guatemala

Michele Monroy-Valle

24. BRAC James P. Grant School of Public Health, Dhaka, Bangladesh

Malay K. Mridha, Kaosar Afsana & Abu A. M. Hanif

25. University of Coimbra, Coimbra, Portugal

Cristina P. Padez, Aristides M. Machado-Rodrigues, Anabela Mota-Pinto, Magdalena Muc, Helena Nogueira & Daniela Rodrigues

26. University of Otago, Dunedin, New Zealand

Andrew Reynolds, Jim I. Mann, Rachael M. McLean, Jody C. Miller & Winsome R. Parnell

27. University of Zagreb, Zagreb, Croatia

Maroje Sorić, Zivka Dika, Bojan Jelaković, Vesna Juresa, Marjeta Majer, Marjeta Mišigoj-Duraković, Sanja Musić Milanović, Vera Musil, Sanja Šalaj, Slavica Sovic & Dorja Vocanec

28. University of Ljubljana, Ljubljana, Slovenia

Maroje Sorić, Gregor Starc, Gregor Jurak, Branimir Leskošek & Vedrana Sember

29. GroundWork, Geneva, Switzerland

James P. Wirth

30. Caja Costarricense de Seguro Social, San José, Costa Rica

Leandra Abarca-Gómez, Alice Bonilla-Vargas, Marvin Cervantes-Loaiza & Roy A. Wong-McClure

31. Al-Quds University, East Jerusalem, State of Palestine

Ziad A. Abdeen & Radwan F. Qasrawi

32. National Center of Public Health, Astana, Kazakhstan

Shynar Abdrakhmanova, Zhamilya Battakova & Tatyana Slazhnyova

33. Qatar University, Doha, Qatar

Hanan F. Abdul Rahim & Zumin Shi

34. Ministry of Health and Social Protection, Dushanbe, Tajikistan

Zulfiya Abdurrahmonova & Sherali Rakhmatulloev

35. Birzeit University, Birzeit, State of Palestine

Niveen M. Abu-Rmeileh & Abdullatif Husseini

36. Usmanu Danfodiyo University Teaching Hospital, Sokoto, Nigeria

Jamila Abubakar Garba

37. Instituto Mexicano del Seguro Social, Mexico City, Mexico

Benjamin Acosta-Cazares, Jorge Escobedo-de la Peña, Margot González-Leon, Ramon A. Rascon-Pacheco & Blanca Sandra Ruiz-Betancourt

38. Qassim University, Unaizah, Saudi Arabia

Ishag Adam

39. RehaKlinika, Rzeszów, Poland

Marzena Adamczyk

40. Flinders University, Adelaide, South Australia, Australia

Robert J. Adams

41. University of Ghana, Accra, Ghana

Seth Adu-Afarwuah & Majid Ezzati

42. University of Copenhagen, Copenhagen, Denmark

Shoaib Afzal, Lars Ängquist, Stig E. Bojesen, Camilla T. Damsgaard, Kim F. Michaelsen, Erik Lykke Mortensen, Børge G. Nordestgaard, Thorkild I. A. Sørensen & Anette Varbo

43. Copenhagen University Hospital, Copenhagen, Denmark

Shoaib Afzal, Stig E. Bojesen, Børge G. Nordestgaard, Peter Schnohr & Anette Varbo

44. University of Oxford, Oxford, UK

Valirie N. Agbor, Zhengming Chen, Timothy Key & Ling Yang

45. Food and Nutrition Research Institute, Taguig, The Philippines

Imelda A. Agdeppa, Mario V. Capanzana, Charmaine A. Duante, Eldridge Ferrer, Glen Gironella & Lynell V. Maniego

46. Urmia University of Medical Sciences, Urmia, Iran

Javad Aghazadeh-Attari & Iraj Mohebbi

47. Ibn Tofail University, Kénitra, Morocco

Hassan Aguenaou

48. Instituto Nacional de Ciencias Médicas y Nutrición, Mexico City, Mexico

Carlos A. Aguilar-Salinas

49. University of Amsterdam, Amsterdam, The Netherlands

Charles Agyemang, Lizzy M. Brewster, Susanne R. de Rooij, Henrike Galenkamp, Karien Stronks, Bert-Jan van den Born, Irene G. M. van Valkengoed & Tanja G. M. Vrijkotte

50. Modeling in Health Research Center, Shahrekord, Iran

Ali Ahmadi

51. Shahid Sadoughi University of Medical Sciences, Yazd, Iran

    Nastaran Ahmadi, Amir Houshang Mehrparvar & Mohammad Reza Mirjalili

52. The Aga Khan University, Karachi, Pakistan

    Imran Ahmed, Sajid Soofi & Zulfiqar A. Bhutta

53. University of Oslo, Oslo, Norway

    Soheir H. Ahmed, Espen Bjertness, Marius B. Bjertness, Aung Soe Htet & Ahmed A. Madar

54. Leibniz Institute for Prevention Research and Epidemiology–BIPS, Bremen, Germany

    Wolfgang Ahrens, Iris Pigeot, Hermann Pohlabeln & Stefan Rach

55. Republican Center for Health Promotion, Bishkek, Kyrgyzstan

    Gulmira Aitmurzaeva & Zhamyila Usupova

56. National Center for Diabetes, Endocrinology and Genetics, Amman, Jordan

    Kamel Ajlouni & Mohammad El-Khateeb

57. Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia

    Hazzaa M. Al-Hazzaa

58. Kuwait Institute for Scientific Research, Kuwait City, Kuwait

    Badreya Al-Lahou

59. King Abdulaziz University, Jeddah, Saudi Arabia

Rajaa Al-Raddadi & Suhad Bahijri

60. The Hashemite University, Zarqa, Jordan

    Huda M. Al Hourani & Buthaina M. Alkhatib

61. Ministry of Health, Kuwait City, Kuwait

    Nawal M. Al Qaoud & Maryam A. Aldwairji

62. Dasman Diabetes Institute, Kuwait City, Kuwait

    Monira Alarouj & Abdullah Alkandari

63. Aldara Hospital and Medical Center, Riyadh, Saudi Arabia

    Fadia AlBuhairan

64. King Abdullah International Medical Research Center, Riyadh, Saudi Arabia

    Shahla AlDhukair

65. Universidade Federal da Integração Latino-Americana, Foz do Iguaçu, Brazil

    Sylvia Alexius

66. Luxembourg Institute of Health, Strassen, Luxembourg

    Ala'a Alkerwi, Gwenaëlle Le Coroller & Maria Ruiz-Castell

67. Barcelona Institute for Global Health CIBERESP, Barcelona, Spain

    Mar Alvarez-Pedrerol, Maribel Casas, Jordi Sunyer & Martine Vrijheid

68. World Health Organization Regional Office for the Eastern Mediterranean, Cairo, Egypt

Eman Aly & Heba M. Fouad

69. Bombay Hospital and Medical Research Centre, Mumbai, India

    Deepak N. Amarapurkar

70. Departamento de Salud del Gobierno Vasco, San Sebastián, Spain

    Pilar Amiano Etxezarreta

71. Ghana Health Service, Kintampo, Ghana

    John Amoah

72. UMR CNRS-MNHN 7206, Paris, France

    Norbert Amougou, Emmanuel Cohen & Patrick Pasquet

73. University of Lille, Lille, France

    Philippe Amouyel, Luc Dauchet, Jonathan Giovannelli & Frederic Gottrand

74. Lille University Hospital, Lille, France

    Philippe Amouyel, Luc Dauchet & Jonathan Giovannelli

75. Western Norway University of Applied Sciences, Sogndal, Norway

    Lars Bo Andersen

76. Norwegian School of Sport Sciences, Oslo, Norway

    Sigmund A. Anderssen, Ulf Ekelund, Elin Kolle & Jostein Steene-Johannessen

77. University of Thessaly, Trikala, Greece

    Odysseas Androutsos

78. Madras Diabetes Research Foundation, Chennai, India

Ranjit Mohan Anjana, Mohan Deepa, Viswanathan Mohan, Rajendra Pradeepa & Coimbatore Subramaniam Shanthirani

79. Zahedan University of Medical Sciences, Zahedan, Iran

Alireza Ansari-Moghaddam & Seyed Mohammad Hashemi-Shahri

80. Yekaterinburg State Medical Academy, Yekaterinburg, Russia

Elena Anufrieva & Olga Petrovna Kovtun

81. National Institute of Public Health, Tunis, Tunisia

Hajer Aounallah-Skhiri & Nada Zoghlami

82. Institute of Public Health of the University of Porto, Porto, Portugal

Joana Araújo & Ana Henriques

83. Norwegian Institute of Public Health, Oslo, Norway

Inger Ariansen, Ingunn Holden Bergh, Sidsel Graff-Iversen, Else Karin Grøholt & Jørgen Meisfjord

84. University of Massachusetts Amherst, Amherst, MA, USA

Raphael E. Arku

85. Public Health Promotion and Development Organization, Kathmandu, Nepal

Krishna K. Aryal & Namuna Shrestha

86. Haramaya University, Dire Dawa, Ethiopia

Nega Aseffa

87. University of Iceland, Reykjavik, Iceland

Thor Aspelund & Vilmundur Gudnason

88. University of Yaoundé 1, Yaoundé, Cameroon

Felix K. Assah, Augustin F. Beybey, François F. Kaze, Jean Claude N. Mbanya, Hermine T. Mopa & Eugène Sobngwi

89. Asfendiyarov Kazakh National Medical University, Almaty, Kazakhstan

Batyrbek Assembekov, Kairat Davletov, Natalya Glushkova & Mukhtar Kulimbet

90. Federal University of Pelotas, Pelotas, Brazil

Maria Cecília F. Assunção, Aluisio J. D. Barros, Paula Duarte de Oliveira, Bruna Gonçalves Cordeiro da Silva, Helen Gonçalves, Bernardo L. Horta, Ana Maria B. Menezes, Isabel O. Oliveira, Ina S. Santos & Cesar G. Victora

91. University of Medicine 1, Yangon, Myanmar

May Soe Aung

92. Oulu University Hospital, Oulu, Finland

Juha Auvinen, Sauli Herrala, Karl-Heinz Herzig, Jari J. Jokelainen & Sirkka Keinänen-Kiukaanniemi

93. University of Oulu, Oulu, Finland

Juha Auvinen, Karl-Heinz Herzig, Marjo-Riitta Jarvelin, Raija Korpelainen & Sylvain Sebert

94. Regional Authority of Public Health, Banska Bystrica, Slovakia

Mária Avdičová & Jana Námešná

95. Tel Aviv University, Tel Aviv, Israel

Shina Avi

96. Hebrew University of Jerusalem, Jerusalem, Israel

Shina Avi

97. University of Porto Medical School, Porto, Portugal

Ana Azevedo & Elisabete Ramos

98. Neyshabur University of Medical Sciences, Neyshabur, Iran

Mohsen Azimi-Nezhad & Seyed Morteza Shamshirgaran

99. Research Institute for Endocrine Sciences, Tehran, Iran

Fereidoun Azizi

100. Indian Council of Medical Research, New Delhi, India

Bontha V. Babu

101. National Institute of Public Health, Copenhagen, Denmark

Maja Bæksgaard Jørgensen & Tina Harmer Lassen

102. National Institute of Pharmacy and Nutrition, Budapest, Hungary

Marta Bakacs & Edit Feigl

103. National Medical Research Centre for Therapy and Preventive Medicine, Moscow, Russia

Yulia Balanova, Vladimir Kutsenko & Svetlana Shalnova

104. University of Science and Technology, Sana'a, Yemen

Mohamed Bamoshmoosh

105. Medical University of Lodz, Lodz, Poland

Maciej Banach, Wojciech Drygas, Elzbieta Dziankowska-Zaborszczyk & Jolanta Slowikowska-Hilczer

106. Universidad Autónoma de Madrid CIBERESP, Madrid, Spain

José R. Banegas, Juan J. Cruz, Pilar Guallar-Castillón, Esther Lopez-Garcia & Fernando Rodríguez-Artalejo

107. University of Rzeszów, Rzeszów, Poland

Joanna Baran, Rafał Baran, Anna Bartosiewicz, Ewelina Czenczek-Lewandowska, Katarzyna Dereń, Justyna Leszczak, Edyta Łuszczki, Piotr Matłosz, Artur Mazur, Grzegorz Sobek, Aneta Weres & Justyna Wyszyńska

108. University of Palermo, Palermo, Italy

Carlo M. Barbagallo & Davide Noto

109. Federal Institute of Education, Science and Technology of Ceara, Ceara, Brazil

Valter Barbosa Filho

110. University of Miami, Miami, FL, USA

Alberto Barceló

111. University Hospital Center Zagreb, Zagreb, Croatia

Maja Baretić, Ana Jelaković, Josipa Josipović, Jelena Kos, Ivan Pećin & Tajana Zeljkovic Vrkic

112. Mohammed V University, Rabat, Morocco

Amina Barkat

113. Unidad de Cirugia Cardiovascular, Guatemala City, Guatemala

Joaquin Barnoya

114. Universidad del Valle, Cali, Colombia

Lena Barrera & Fabián Méndez

115. National Institute of Health Doutor Ricardo Jorge, Lisbon, Portugal

Marta Barreto, Baltazar Nunes & Ana I. Rito

116. NOVA University Lisbon, Lisbon, Portugal

Marta Barreto & Baltazar Nunes

117. University of Pernambuco, Recife, Brazil

Mauro Virgílio Gomes Barros, Jorge Bezerra, Marcos André Moura-dos-Santos, Caroline Ramos de Moura Silva, Fernanda Cunha Soares & Rildo de Souza Wanderley Júnior

118. Baqai Institute of Diabetology and Endocrinology, Karachi, Pakistan

Abdul Basit

119. Federal University of Santa Catarina, Florianópolis, Brazil

Joao Luiz D. Bastos, Eleonora d'Orsi, Maria Alice Altenburg de Assis, Patrícia de Fragas Hinnig, Thamara Hubler Figueiró, Marcus V. V. Lopes, Karina Mary de Paiva, Cassiano Ricardo Rech, Diego Augusto Santos Silva & Kelly S. Silva

120. Dalhousie University, Halifax, Nova Scotia, Canada

Iqbal Bata & Ronald D. Gregor

121. Jordan University of Science and Technology, Irbid, Jordan

Anwar M. Batieha & Yousef Saleh Khader

122. Universidade Federal de Ouro Preto, Ouro Preto, Brazil

Aline P. Batista & George L. L. Machado-Coelho

123. Federal University of Maranhão, São Luís, Brazil

Rosangela L. Batista, Susana C. Confortin & Antonio M. Silva

124. University of Sydney, Sydney, New South Wales, Australia

Louise A. Baur, Chris Cowell, Sarah P. Garnett & Alison J. Hayes

125. Cliniques Universitaires de Kinshasa, Kinshasa, Democratic Republic of the Congo

Pascal M. Bayauli

126. University of Auckland, Auckland, New Zealand

Robert Beaglehole, Rod T. Jackson, Jacqueline Ramke & Boyd A. Swinburn

127. University College Dublin, Dublin, Ireland

Silvia Bel-Serrat, Cecily Kelleher & Breige A. McNulty

128. Christian Medical College, Vellore, India

Antonisamy Belavendra, Gowri Mahasampath & Nihal Thomas

129. University Tunis El Manar, Tunis, Tunisia

Habiba Ben Romdhane & Olfa Saidi

130. Federal Ministry of Social Affairs, Health, Care and Consumer Protection, Vienna, Austria

Judith Benedics & Adelheid Weber

131. Cafam University Foundation, Bogotá, Colombia

Mikhail Benet

132. Ministerio de Salud Pública y Bienestar Social, Asunción, Paraguay

Gilda Estela Benitez Rolandi, Felicia Cañete, Ethel Santacruz & Guillermo Sequera

133. University of Agder, Kristiansand, Norway

Elling Bere & Frøydis N. Vik

134. Addis Continental Institute of Public Health, Addis Ababa, Ethiopia

Yemane Berhane

135. Kazakh National Medical University, Almaty, Kazakhstan

Salim Berkinbayev & Ildar Fakhradiyev

136. Universidad Peruana Cayetano Heredia, Lima, Peru

Antonio Bernabe-Ortiz, Maria Lazo-Porras, J. Jaime Miranda, Salim Mohanna, Pedro J. Ortiz & Tania Tello

137. Lithuanian University of Health Sciences, Kaunas, Lithuania

Gailute Bernotiene, Virginija Dulskiene, Jurate Klumbiene, Vilma Kriaucioniene, Renata Kuciene, Dalia Luksiene, Jurate Medzioniene, Janina Petkeviciene, Ausra Petrauskiene, Ricardas Radisauskas, Abdonas Tamosiunas & Justina Vaitkeviciute

138. Pontificia Universidad Católica de Chile, Santiago, Chile

Ximena Berrios Carrasola, Guadalupe Echeverría, Catterina Ferreccio, Paula Margozzini, Juan Francisco Miquel, Flavio

Nervi, Attilio Rigotti & Gonzalo Valdivia

139. University of São Paulo, São Paulo, Brazil

Heloísa Bettiol, Viviane C. Cardoso, Andrea R. V. R. Horimoto, Paulo A. Lotufo, Alicia Matijasevich & Alexandre C. Pereira

140. Johannes Gutenberg University, Mainz, Germany

Manfred E. Beutel, Karl J. Lackner, Martina Müller-Nurasyid, Thomas Münzel, Norbert Pfeiffer & Philipp S. Wild

141. B. J. Medical College, Ahmedabad, India

Aroor Bhagyalaxmi

142. Chirayu Medical College, New Delhi, India

Sumit Bharadwaj

143. Sunder Lal Jain Hospital, Delhi, India

Santosh K. Bhargava

144. Shandong University of Traditional Chinese Medicine, Jinan, China

Hongsheng Bi & Jianfeng Wu

145. Shanghai Jiao-Tong University School of Medicine, Shanghai, China

Yufang Bi & Guang Ning

146. Universidad de la República, Montevideo, Uruguay

Daniel Bia, Jose G. Boggia, Oscar A. Noboa & Yanina Zocalo

147. University of Strasbourg, Strasbourg, France

Katia Biasch & Marie Moitry

148. Institute of Medical Research and Medicinal Plant Studies, Yaoundé, Cameroon

Elysée Claude Bika Lele

149. Ufa Eye Research Institute, Ufa, Russia

Mukharram M. Bikbov, Albina A. Fakhretdinova, Ellina M. Iakupova, Gyulli M. Kazakbaeva, Iuliia A. Rusakova & Azaliia M. Tuliakova

150. Nepal Health Research Council, Kathmandu, Nepal

Bihungum Bista, Meghnath Dhimal & Anil Poudyal

151. University of Montenegro, Niksic, Montenegro

Dusko J. Bjelica & Stevo R. Popovic

152. University of Southern Denmark, Copenhagen, Denmark

Peter Bjerregaard, Louise Eriksen & Janne S. Tolstrup

153. University of Gothenburg, Gothenburg, Sweden

Cecilia Björkelund, Robert Eggertsen, Dominique Hange, Jenny M. Kindblom, Lauren Lissner, Kirsten Mehlig, Claes Ohlsson, Annika Rosengren & Agneta Sjöberg

154. Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil

Katia V. Bloch

155. National Institute for Public Health and the Environment, Bilthoven, The Netherlands

Anneke Blokstra, Jolanda M. A. Boer, Johannes Brug, W. M. Monique Verschuren & Lucie Viet

156. University of Haifa, Haifa, Israel

Moran Blychfeld Magnazu & Ronit Endevelt

157. Ministry of Health, Ramat Gan, Israel

Moran Blychfeld Magnazu

158. University of Turin, Turin, Italy

Simona Bo, Graziella Bruno & Gabriella Gruden

159. University College London, London, UK

Martin Bobak, Nish Chaturvedi, Hynek Pikhart, Rui Providencia, Reecha Sofat, Julie Taylor, S. Goya Wannamethee & Andrew Wong

160. Liverpool John Moores University, Liverpool, UK

Lynne M. Boddy

161. Nanyang Technological University, Singapore, Singapore

Bernhard O. Boehm

162. National Medical Research Center for Endocrinology, Moscow, Russia

Elena Bogova & Valentina Peterkova

163. Centro de Educación Médica e Investigaciones Clínicas, Buenos Aires, Argentina

Carlos P. Boissonnet

164. IRCCS Neuromed, Pozzilli, Italy

Marialaura Bonaccio, Simona Costanzo, Amalia De Curtis, Giovanni de Gaetano, Maria Benedetta Donati & Licia Iacoviello

165. Toulouse University School of Medicine, Toulouse, France

Vanina Bongard, Jean Ferrieres & Jean-Bernard Ruidavets

166. University of Zurich, Zurich, Switzerland

Matthias Bopp, David Faeh, Felix Gutzwiller, Susi Kriemler, Frank J. Rühli & Kaspar Staub

167. KU Leuven, Leuven, Belgium

Herman Borghs, Frank Buntinx, Frank Claessens, Jesus D. Melgarejo, Jan A. Staessen, Dirk Vanderschueren, Yu-Ling Yu & Zhen-Yu Zhang

168. Ministry of Health, Victoria, Seychelles

Pascal Bovet & Bharathi Viswanathan

169. Unisanté, Lausanne, Switzerland

Pascal Bovet & Fred Michel Paccaud

170. World Health Organization Country Office in Tajikistan, Dushanbe, Tajikistan

Khadichamo Boymatova

171. Flemish Agency for Care and Health, Brussels, Belgium

Lien Braeckevelt

172. Ghent University, Ghent, Belgium

Lutgart Braeckman, Els Clays, Guy De Backer, Dirk De Bacquer, Stefaan De Henauw, Delphine De Smedt, Patrick Kolsteren, Carl Lachat, Nathalie Michels, Koen Van Herck, Wendy Van Lippevelde & Maïté Verloigne

173. FrieslandCampina, Amersfoort, The Netherlands

Marjolijn C. E. Bragt & Ilse M. S. L. Khouw

174. Universidad Central de Venezuela, Caracas, Venezuela

Imperia Brajkovich, Marianella Herrera-Cuenca & Maritza Landaeta Jimenez

175. Bielefeld University, Bielefeld, Germany

Juergen Breckenkamp

176. World Health Organization Athens Quality of Care Office, Athens, Greece

João Breda

177. German Cancer Research Center, Heidelberg, Germany

Hermann Brenner, Rudolf Kaaks, Kai-Uwe Saum & Ben Schöttker

178. The Fred Hollows Foundation, Auckland, New Zealand

Garry R. Brian

179. University of the Andes, Mérida, Venezuela

Yajaira Briceño & Mariela Paoli

180. Carol Davila University of Medicine and Pharmacy, Bucharest, Romania

Lacramioara Brinduse, Eliza Cinteza, Maria Dorobantu & Oana-Florentina Gheorghe-Fronea

181. Instituto Politécnico de Lisboa, Lisbon, Portugal

Miguel Brito

182. Swansea University, Swansea, UK

Sinead Brophy & Gareth Stratton

183. University College Copenhagen, Copenhagen, Denmark

Anna Bugge

184. Institute of Public Health, Tirana, Albania

Genc Burazeri & Jolanda Hyska

185. Munster Technological University, Cork, Ireland

Con Burns, Tara C. Coppinger & Janette Walton

186. Universidad de La Laguna, Santa Cruz de Tenerife, Spain

Antonio Cabrera de León

187. University of Malta, Msida, Malta

Joseph Cacciottolo & Sarah Cuschieri

188. Vanderbilt University, Nashville, TN, USA

Hui Cai & Wei Zheng

189. Pan American Health Organization, Washington, DC, USA

Roberta B. Caixeta & Pedro Ordunez

190. Ministry of Health, Tongatapu, Tonga

Tilema Cama

191. Canadian Fitness and Lifestyle Research Institute, Ottawa, Ontario, Canada

Christine Cameron & Cora L. Craig

192. Hospital Santa Maria, Lisbon, Portugal

José Camolas

193. Istanbul University–Cerrahpasa, Istanbul, Turkey

Günay Can

194. Universidade Federal de Juiz de Fora, Juiz de Fora, Brazil

Ana Paula C. Cândido

195. National Institute of Public Health, Prague, Czech Republic

Naděžda Čapková, Jana Kratenova, Ruzena Kubinova, Michala Lustigová & Kristyna Zejglicova

196. Gaetano Fucito Hospital, Mercato San Severino, Italy

Eduardo Capuano, Rocco Capuano & Vincenzo Capuano

197. University of Groningen, Groningen, The Netherlands

Marloes Cardol, Eva Corpeleijn & Daan Kromhout

198. Karolinska Institutet, Huddinge, Sweden

Axel C. Carlsson & Holger Theobald

199. Centro de Estudios Sobre Nutrición Infantil, Buenos Aires, Argentina

Esteban Carmuega & Maria Elisa Zapata

200. University of Porto, Porto, Portugal

Joana Carvalho, Luis Lopes, Nuno Lunet, Jorge Mota, Andreia N. Pizarro, Maria Paula Santos & Rute Santos

201. University of Zaragoza, Zaragoza, Spain

José A. Casajús, Pilar De Miguel-Etayo, Esther M. González-Gil & Luis A. Moreno

202. Santiago de Compostela University, Santiago de Compostela, Spain

Felipe F. Casanueva

203. Ministry of Health, Ankara, Turkey

Ertugrul Celikcan, Sibel Gogen & Nazan Yardim

204. Council for Agricultural Research and Economics, Rome, Italy

Laura Censi, Marika Ferrari, Myriam Galfo & Romana Roccaldo

205. Federal University of Rio Grande, Rio Grande, Brazil

Juraci A. Cesar & Samuel C. Dumith

206. India Diabetes Research Foundation, Chennai, India

Snehalatha Chamukuttan, Ambady Ramachandran & Mary Simon

207. Duke–NUS Medical School, Singapore, Singapore

Angelique Chan, Tazeen Jafar, Rahul Malhotra & Tien Yin Wong

208. ICMR–National Institute of Medical Statistics, New Delhi, India

Himanshu K. Chaturvedi, Geetha R. Menon & Arvind Pandey

209. Singapore Eye Research Institute, Singapore, Singapore

Miao Li Chee, Charumathi Sabanayagam, Rosalynn Siantar & Yih Chung Tham

210. Academia Sinica, Taipei, Taiwan

Chien-Jen Chen & Wen-Harn Pan

211. Capital Institute of Pediatrics, Beijing, China

Fangfang Chen & Jie Mi

212. Xiangtan University, Xiangtan, China

Huashuai Chen

213. Kailuan General Hospital, Tangshan, China

Shuohua Chen, Liufu Cui, Jingli Gao & Shouling Wu

214. National University of Singapore, Singapore, Singapore

Ching-Yu Cheng, Xinyi Gwee, Jeannette Lee, Charlie Lim, Wei-Yen Lim, Ei Ei K. Nang, Tze Pin Ng, Xueling Sim & E. Shyong Tai

215. US Centers for Disease Control and Prevention, Atlanta, GA, USA

Yiling J. Cheng

216. Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran

Bahman Cheraghian, Mohammad Esmaeel Motlagh, Nader Saki & Ali Akbar Shayesteh

217. The Gertner Institute for Epidemiology and Health Policy Research, Ramat Gan, Israel

Angela Chetrit, Rachel Dankner, Ofra Kalter-Leibovici & Gilad Twig

218. National Centre of Public Health and Analyses, Sofia, Bulgaria

Ekaterina Chikova-Iscener, Vesselka L. Duleva & Lalka S. Rangelova

219. Amsterdam UMC Public Health Research Institute, Amsterdam, The Netherlands

Mai J. M. Chinapaw & Han C. G. Kemper

220. Universidad de Costa Rica, San José, Costa Rica

Anne Chinnock & Georgina Gómez

221. University of Fribourg, Fribourg, Switzerland

Arnaud Chiolero

222. National Taiwan University, Taipei, Taiwan

Shu-Ti Chiou & Hsien-Ho Lin

223. CIBERESP, Madrid, Spain

María-Dolores Chirlaque, Manoli Garcia-de-la-Hera, Magda Gasull, José María Huerta, Jesús M. Ibarluzea, Diego Salmerón, Helmut Schröder & Laura Torres-Collado

224. Seoul National University, Seoul, Republic of Korea

Belong Cho & Jin Soo Moon

225. University of Southern Denmark, Odense, Denmark

Kaare Christensen, Anders Grøntved, Peter Lund Kristensen, Niels C. Møller, Stine Schramm, Morten Sodemann & Niels Wedderkopp

226. Universidade Estadual Paulista, Presidente Prudente, Brazil

Diego G. Christofaro, Romulo A. Fernandes & William R. Tebar

227. Medical University of Silesia, Katowice, Poland

Jerzy Chudek & Andrzej Więcek

228. Charles University, Prague, Czech Republic

Renata Cifkova & Michala Lustigová

229. Thomayer Hospital, Prague, Czech Republic

Renata Cifkova

230. Primary Health Care, Floriana, Malta

Michelle Cilia, Victoria Farrugia Sant'Angelo & Lorraine Tabone

231. University of Salerno, Fisciano, Italy

Massimo Cirillo

232. Statistics Canada, Ottawa, Ontario, Canada

Janine Clarke, Thomas Ferrao, Anie Marcil, Scott B. McLean, Joel G. R. Roy & Jennifer Servais

233. Alicante Institute for Health and Biomedical Research, Alicante, Spain

Laura-María Compañ-Gabucio

234. Agency for Preventive and Social Medicine, Bregenz, Austria

Hans Concin, Harald Geiger, Georg Posch & Emanuel Zitt

235. University of Southampton, Southampton, UK

Cyrus Cooper, Elaine Dennison, Caroline H. Fall, Allan G. Hill, Clive Osmond, Senthil K. Vasan & Leo D. Westbury

236. Pontificia Universidad Javeriana, Bogotá, Colombia

Lilia Yadira Cortés, Luis F. Gomez & Luz Nayibe Vargas

237. Institut Pasteur de Lille, Lille, France

Dominique Cottel & Jean Dallongeville

238. Malawi Epidemiology and Intervention Research Unit, Lilongwe, Malawi

Amelia C. Crampin

239. CIBEROBN, Madrid, Spain

Ana B. Crujeiras, Pilar De Miguel-Etayo, Luis A. Moreno & Dora Romaguera

240. Hungarian University of Sports Science, Budapest, Hungary

Tamás Csányi

241. University of Debrecen, Debrecen, Hungary

Semánová Csilla

242. National Institute of Public Health, Bucharest, Romania

Alexandra M. Cucu, Constanta Huidumac Petrescu & Rodica Nicolescu

243. University of Medicine and Pharmacy, Bucharest, Romania

Alexandra M. Cucu

244. Universidade Federal do Rio Grande do Norte, Natal, Brazil

Felipe V. Cureau

245. National Research Council, Reggio Calabria, Italy

Graziella D'Arrigo

246. Eftimie Murgu University Resita, Resita, Romania

Liliana Dacica

247. Eduardo Mondlane University, Maputo, Mozambique

Albertino Damasceno

248. Indian Statistical Institute, Kolkata, India

Parasmani Dasgupta

249. Tabriz Health Services Management Research Center, Tabriz, Iran

Saeed Dastgiri

250. Geneva University Hospitals, Geneva, Switzerland

David De Ridder & Idris Guessous

251. Sciensano, Brussels, Belgium

Karin De Ridder, Stefaan Demarest & Johan Van der Heyden

252. University Medical Centers, Amsterdam, The Netherlands

Susanne R. de Rooij & Tanja G. M. Vrijkotte

253. National Research Centre for Preventive Medicine, Moscow, Russia

Alexander D. Deev

254. Innovating Health International, Port-au-Prince, Haiti

Vincent DeGennaro Jr

255. University of Montreal, Montreal, Quebec, Canada

Hélène Delisle

256. French National Research Institute for Sustainable Development, Montpellier, France

Francis Delpeuch, Michelle Holdsworth, Yves Kameli, Edwige Landais, Agnès Le Port, Bernard Maire, Yves Martin-Prevel, Mathilde Savy & Pierre Traissac

257. French Public Health Agency, St Maurice, France

Valérie Deschamps & Benoit Salanave

258. Mediterranea Cardiocentro, Naples, Italy

Augusto Di Castelnuovo, Francesco Gianfagna & Sara Magnacca

259. Universidade do Vale do Rio dos Sinos, São Leopoldo, Brazil

Juvenal Soares Dias-da-Costa & Marcos Pattussi

260. National Institute of Hygiene, Epidemiology and Microbiology, Havana, Cuba

María Elena Díaz-Sánchez, Ramón Suárez-Medina & Patricia Varona-Pérez

261. National Council of Scientific and Technical Research, Buenos Aires, Argentina

Alejandro Diaz

262. Servicio Canario de la Salud del Gobierno de Canarias, Santa Cruz de Tenerife, Spain

Pedro Díaz Fernández, María José García Mérida & María del Cristo Rodriguez-Perez

263. Consejería de Salud del Gobierno de La Rioja, Logroño, Spain

María Pilar Díez Ripollés

264. Ministry of Health and Medical Education, Tehran, Iran

Shirin Djalalinia & Niloofar Peykari

265. University of Novi Sad, Novi Sad, Serbia

Visnja Djordjic, Sergej M. Ostojic & Ivana Radic

266. National Institute of Nutrition, Hanoi, Vietnam

Ha T. P. Do, Chinh Nguyen Huu, Khanh Le Nguyen Bao & Tuyen D. Le

267. University of Queensland, Brisbane, Queensland, Australia

Annette J. Dobson

268. Instituto de Investigación Nutricional, Lima, Peru

Liria Dominguez & Rossina G. Pareja

269. Istituto Superiore di Sanità, Rome, Italy

Chiara Donfrancesco, Simona Giampaoli, Laura Lauria, Paola Nardone, Luigi Palmieri, Daniela Pierannunzio & Angela Spinelli

270. Sun Yat-sen University, Guangzhou, China

Guanghui Dong & Yingfeng Zheng

271. Universidad de Cuenca, Cuenca, Ecuador

Silvana P. Donoso & Angélica M. Ochoa-Avilés

272. Helmholtz Zentrum München, Munich, Germany

Angela Döring, Christian Gieger, Margit Heier, Birgit Linkohr, Christa Meisinger, Annette Peters, Andrea Schneider, Jutta Stieber, Doris Stöckl, Barbara Thorand & Kathrin Wolf

273. Tehran University of Medical Sciences, Tehran, Iran

Ahmad Reza Dorosty, Sareh Eghtesad, Fatemeh Malekzadeh, Reza Malekzadeh, Kazem Mohammad, Mohammad Reza Mohammadi, Zahra Mohammadi, Seyed-Ali Mostafavi, Mahboubeh Parsaeian, Hamed Pouraram, Akram Pourshams, Hossein Poustchi, Alireza Sadjadi, Sadaf G. Sepanlou, Maryam Sharafkhah & Amaneh Shayanrad

274. Ministère de la Santé et de l’Hygiène Publique, Abidjan, Côte d’Ivoire

Kouamelan Doua

275. University Hospital Düsseldorf, Düsseldorf, Germany

Nico Dragano

276. National Institute of Cardiology, Warsaw, Poland

Wojciech Drygas, Pawel Kurjata & Aleksandra Piwonska

277. Beijing Center for Disease Prevention and Control, Beijing, China

Jia Li Duan

278. IRL 3189 ESS, Marseille, France

Priscilla Duboz & Enguerran Macia

279. Scuola Superiore Sant’Anna, Pisa, Italy

Anar Dushpanova

280. Al-Farabi Kazakh National University, Almaty, Kazakhstan

Anar Dushpanova, Natalya Glushkova, Zhanna Kalmatayeva & Mukhtar Kulimbet

281. Semey Medical University, Semey, Kazakhstan

Azhar Dyussupova

282. University of Latvia, Riga, Latvia

Vilnis Dzerve & Elena Pahomova

283. Ministry of Health and Medical Services, Gizo, Solomon Islands

Ricky Eddie

284. Hormozgan University of Medical Sciences, Bandar Abbas, Iran

Ebrahim Eftekhar & Azim Nejatizadeh

285. University of Benin, Benin City, Nigeria

Eruke E. Egbagbe

286. University of Skövde, Skövde, Sweden

Gabriele Eiben

287. Ministry of Health, Rabat, Morocco

Laila El Ammari

288. National Institute of Nutrition and Food Technology, Tunis, Tunisia

Jalila El Ati

289. The University of the West Indies, Kingston, Jamaica

Denise Eldemire-Shearer, Trevor S. Ferguson, Damian K. Francis, Kenneth James, Shelly R. McFarlane, Marshall K. Tulloch-Reid, Rainford J. Wilks & Novie O. Younger-Coleman

290. Ministry of Health, Jerusalem, Israel

Ronit Endevelt

291. University of California Davis, Davis, CA, USA

Reina Engle-Stone

292. University of Stellenbosch, Cape Town, South Africa

Rajiv T. Erasmus

293. University of Duisburg-Essen, Essen, Germany

Raimund Erbel, Karl-Heinz Jöckel, Nils Lehmann, Ulla Roggenbuck & Sara Schramm

294. Karadeniz Technical University, Trabzon, Turkey

Cihangir Erem, Irfan Nuhoğlu & Murat Topbas

295. Dokuz Eylul University, Izmir, Turkey

Gul Ergor & Belgin Unal

296. University of Helsinki, Helsinki, Finland

Johan G. Eriksson & Jukka T. Salonen

297. Mashhad University of Medical Sciences, Mashhad, Iran

Saeid Eslami, Maryam Keramati, Reza Mohammadpourhodki & Seyed Rasoul Zakavi

298. Rafsanjan University of Medical Sciences, Rafsanjan, Iran

Ali Esmaeili & Hamid Hakimi

299. Queen's University Belfast, Belfast, UK

Alun Evans & Dermot O'Reilly

300. Tabriz University of Medical Sciences, Tabriz, Iran

Elnaz Faramarzi, Saeid Safiri & Mohammad Hossein Somi

301. Fasa University of Medical Sciences, Fasa, Iran

Mojtaba Farjam

302. Shiraz University of Medical Sciences, Shiraz, Iran

Mohammad Reza Fattahi, Reza Malekzadeh, Salar Rahimikazerooni, Abbas Rezaianzadeh & Ali Reza Safarpour

303. Baqai Medical University, Karachi, Pakistan

Asher Fawwad

304. Centro de Salud Villanueva Norte, Badajoz, Spain

Francisco J. Felix-Redondo

305. Hospital Don Benito-Villanueva de la Serena, Badajoz, Spain

Daniel Fernández-Bergés

306. Ministry of Health, Buenos Aires, Argentina

Daniel Ferrante

307. Universidad de Santiago de Chile, Santiago, Chile

Gerson Ferrari

308. University of Insubria, Varese, Italy

Marco M. Ferrario, Alessandro Gialluisi, Francesco Gianfagna, Licia Iacoviello & Giovanni Veronesi

309. Federal University of Alagoas, Alagoas, Brazil

Haroldo S. Ferreira

310. Institute of Mother and Child, Warsaw, Poland

Anna Fijalkowska, Magdalena Korzycka & Hanna Nalecz

311. Hospital Infantil Sabará, São Paulo, Brazil

Mauro Fisberg

312. University of Tartu, Tartu, Estonia

Krista Fischer, Kairit Mikkel & Mari-Liis Tammesoo

313. Universiti Sains Malaysia, Kelantan, Malaysia

Leng Huat Foo, Kamarul Imran Musa, Wan Mohamad Wan Bebakar & Nor Azwany Yaacob

314. Umeå University, Umeå, Sweden

Maria Forsner, Maria Nordendahl & Stefan Söderberg

315. Federal University of São Paulo, São Paulo, Brazil

Maria do Carmo Franco & Maria Wany Strufaldi

316. University Clinical Centre Ljubljana, Ljubljana, Slovenia

Zlatko Fras

317. Hospital Universitario Son Espases, Palma, Spain

Guillermo Frontera

318. Hospital de Clinicas de Porto Alegre, Porto Alegre, Brazil

Flavio D. Fuchs

319. Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil

Sandra C. Fuchs, Adroaldo Cesar Araujo Gaya, Anelise Reis Gaya, João Mello, Leila B. Moreira, Maria Teresa A. Olinto, Mariana

Sbaraini, Beatriz D. Schaan, Karen Sparrenberger & Daniel G. Vasques

320. Universitas Sumatera Utara, Medan, Indonesia

Isti I. Fujiati

321. Kindai University, Osaka-Sayama, Japan

Yuki Fujita

322. Kyoto University, Kyoto, Japan

Matsuda Fumihiko, Takuro Furusawa & Tabara Yasuharu

323. I. Horbachevsky Ternopil National Medical University, Ternopil, Ukraine

Viktoriya Furdela & Halyna Pavlyshyn

324. Medical University of Warsaw, Warsaw, Poland

Zbigniew Gaciong

325. Consejería de Sanidad del Gobierno de Cantabria, Santander, Spain

Manuel Galán Cuesta

326. Jagiellonian University Medical College, Kraków, Poland

Andrzej Galbarczyk, Grazyna Jasienska, Magdalena Klimek, Karolina Miłkowska, Ilona Nenko, Andrzej Pająk & Roman Topór-Mądry

327. Ministero della Salute DG Prevenzione Sanitaria, Rome, Italy

Daniela Galeone & Maria Teresa Menzano

328. University of Catania, Catania, Italy

Fabio Galvano, Justyna Godos, Giuseppe Grosso, Stefano Marventano & Antonio Mistretta

329. Agencia Española de Seguridad Alimentaria y Nutrición, Madrid, Spain

Marta García Solano & Enrique Gutiérrez González

330. Africa Health Research Institute, Mtubatuba, South Africa

Dickman Gareta

331. Geneva University Medical School, Geneva, Switzerland

Jean-Michel Gaspoz

332. Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

Andrea Gazzinelli & Gustavo Velasquez-Melendez

333. Utrecht University, Utrecht, The Netherlands

Ulrike Gehring, N. Charlotte Onland-Moret & Yvonne T. van der Schouw

334. Wageningen University, Wageningen, The Netherlands

Johanna M. Geleijnse

335. Medical Research Foundation, Chennai, India

Ronnie George & Vijaya Lingam

336. Kurdistan University of Medical Sciences, Sanandaj, Iran

Ebrahim Ghaderi & Farhad Moradpour

337. University of Adelaide, Adelaide, South Australia, Australia

Tiffany K. Gill, David A. Gonzalez-Chica, Konrad Jamrozik & Anne Taylor

338. Lund University, Lund, Sweden

Aleksander Giwercman & Emily Sonestedt

339. Aristotle University of Thessaloniki, Thessaloniki, Greece

Konstantinos Gkiouras, Xenophon Theodoridis & Areti Triantafyllou

340. National Institute for Health Development, Tallinn, Estonia

Natalja Gluškova, Merike Liivak, Helle-Mai Loit, Keiu Nelis, Eha Nurk, Marge Saamel & Toomas Veidebaum

341. Central University of Kerala, Kasaragod, India

Ramesh Godara & K. R. Thankappan

342. Institut National de la Santé et de la Recherche Médicale, Paris, France

Marcel Goldberg & Marie Zins

343. Paris University, Paris, France

Marcel Goldberg & Marie Zins

344. Instituto Murciano de Investigación Biosanitaria Virgen de la Arrixaca, Murcia, Spain

Jesús Humberto Gómez Gómez & Inés Sánchez Rodríguez

345. Gasol Foundation, Sant Boi de Llobregat, Spain

Santiago F. Gómez

346. University of Lleida, Sant Boi de Llobregat, Spain

Santiago F. Gómez

347. PASs Hirszfeld Institute of Immunology and Experimental Therapy, Wroclaw, Poland

Aleksandra Gomula, Slawomir Koziel & Natalia Nowak-Szczepanska

348. University Agostinho Neto, Luanda, Angola

Mauer Gonçalves

349. Kansas State University, Manhattan, KS, USA

Ana D. González-Alvarez

350. Universidad Politécnica de Madrid, Madrid, Spain

Marcela Gonzalez-Gross

351. International Clinical Research Center, Brno, Czech Republic

Juan P. González-Rivas

352. National Institute of Public Health, Cuernavaca, Mexico

Clicerio González-Villalpando, Leticia Hernandez Cadena, Eric Monterrubio-Flores, Juan A. Rivera, Rosalba Rojas-Martinez, Eduardo Salazar Martinez, Teresa Shamah-Levy & Salvador Villalpando

353. Centro de Estudios en Diabetes A.C., Mexico City, Mexico

María-Elena González-Villalpando

354. Universidad Autónoma de Santo Domingo, Santo Domingo, Dominican Republic

Angel R. Gonzalez

355. Ministry of Health, Lisbon, Portugal

Antonio Pedro Graça & Maria João Gregório

356. Institute for Clinical and Experimental Medicine, Prague, Czech Republic

Dušan Grafnetter & Vera Lanska

357. Children's Memorial Health Institute, Warsaw, Poland

Aneta Grajda, Zbigniew Kulaga & Mieczyslaw Litwin

358. University of Thessaly, Larissa, Greece

Maria G. Grammatikopoulou

359. National Center of Cardiovascular Diseases, Beijing, China

Dongfeng Gu

360. International Life Science Institute, Buenos Aires, Argentina

Viviana Guajardo

361. University of Ferrara, Ferrara, Italy

Emanuela Gualdi-Russo, Natascia Rinaldo & Luciana Zaccagni

362. Authority Sanitaria San Marino, San Marino, San Marino

Andrea Gualtieri & Elena Sacchini

363. Icelandic Heart Association, Kopavogur, Iceland

Elias F. Gudmundsson, Johanna Gunnlaugsdottir & Örn Olafsson

364. Universidad Icesi, Cali, Colombia

Ramiro Guerrero & Victoria E. Soto-Rojas

365. State University of Montes Claros, Montes Claros, Brazil

Andre L. Guimaraes

366. King's College London, London, UK

Martin C. Gulliford & Kenji Shibuya

367. International Agency for Research on Cancer, Lyon, France

Marc J. Gunter & Inge Huybrechts

368. Capital Medical University, Beijing, China

Xiu-Hua Guo, Ya Xing Wang, Wenbin Wei, Li Juan Wu & Qi Sheng You

369. Capital Medical University Beijing Tongren Hospital, Beijing, China

Yin Guo, Jie Hao, Lijuan Liu & Ningli Wang

370. Healis-Sekhsaria Institute for Public Health, Navi Mumbai, India

Prakash C. Gupta, Sameer Narake & Mangesh S. Pednekar

371. Eternal Heart Care Centre and Research Institute, Jaipur, India

Rajeev Gupta

372. University of Ibadan, Ibadan, Nigeria

Oye Gureje

373. Institute for Clinical Effectiveness and Health Policy, Buenos Aires, Argentina

Laura Gutierrez, Vilma E. Irazola & Adolfo Rubinstein

374. National Health Insurance Service, Wonju, Republic of Korea

Seongjun Ha & Yeunsook Rho

375. Prevention of Metabolic Disorders Research Center, Tehran, Iran

Farzad Hadaegh

376. Research and Education Institute of Child Health, Nicosia, Cyprus

Charalambos A. Hadjigeorgiou, Savvas Savva & Michael J. Tornaritis

377. Danish Cancer Society Research Center, Copenhagen, Denmark

Jytte Halkjær & Anne Tjonneland

378. The University of the West Indies, Cave Hill, Barbados

Ian R. Hambleton & Christina Howitt

379. Kermanshah University of Medical Sciences, Kermanshah, Iran

Behrooz Hamzeh & Farid Najafi

380. Africa Health Research Institute, Durban, South Africa

Willem A. Hanekom & Emily B. Wong

381. Federal University of Pernambuco, Recife, Brazil

Carla Menêses Hardman & Rafael dos Santos Henrique

382. Yasuj University of Medical Sciences, Yasuj, Iran

Javad Harooni & Moslem Sedaghattalab

383. International Hellenic University, Thessaloniki, Greece

Maria Hassapidou, Ioannis Pagkalos, Sousana K. Papadopoulou, Dimitrios Poulimeneas, Maria Tsigga & Panagiotis Tsintavis

384. Kyushu University, Fukuoka, Japan

Jun Hata & Toshiharu Ninomiya

385. University of Bergen, Bergen, Norway

Teresa Haugsgjerd & Grethe S. Tell

386. Tulane University, New Orleans, LA, USA

Jiang He

387. National Research Institute for Health and Family Planning, Beijing, China

Yuan He & Xu Ma

388. Chinese Center for Disease Control and Prevention, Beijing, China

Yuna He, Xiaoguang Yang, Wenhua Zhao & Maigeng Zhou

389. University of Pécs, Pécs, Hungary

Regina Heidinger-Felső, Dénes Molnár & Judit A. Repasy

390. Danish Health Authority, Copenhagen, Denmark

Tatjana Hejgaard

391. Joep Lange Institute, Amsterdam, The Netherlands

Marleen Elisabeth Hendriks

392. Universidad Autónoma de Bucaramanga, Bucaramanga, Colombia

Victor M. Herrera

393. ETH Zurich, Zurich, Switzerland

Isabelle Herter-Aeberli

394. Chronic Diseases Research Center, Tehran, Iran

Ramin Heshmat

395. University of Hong Kong, Hong Kong, China

Sai Yin Ho, Tai Hing Lam, Gabriel M. Leung, Michael Y. Ni & Hung-Kwan So

396. The Chinese University of Hong Kong, Hong Kong, China

Suzanne C. Ho, Yn-Tz Sung & Jean Woo

397. University of Western Australia, Perth, Western Australia, Australia

Michael Hobbs, Debra S. Judge & Phoebe R. Spencer

398. Universidade Federal do Paraná, Curitiba, Brazil

Doroteia A. Höfelmann & Lariane M. Ono

399. Shahid Beheshti University of Medical Sciences, Tehran, Iran

Reza Homayounfar, Davood Khalili, Bahareh Kheiri, Amirabbas Momenan, Hamideh Sabbaghi & Sare Safi

400. Gasol Foundation, Barcelona, Spain

Clara Homs

401. University Ramon Llull, Sant Boi de Llobregat, Spain

Clara Homs

402. Kingston Health Sciences Centre, Kingston, Ontario, Canada

Wilma M. Hopman

403. Fundación Oftalmológica de Santander, Bucaramanga, Colombia

Claudia M. Hormiga

404. University Oran 1, Oran, Algeria

Leila Houti & Sounnia Mediene Benchekor

405. Independent Public Health Specialist, Nay Pyi Taw, Myanmar

Thein Thein Htay

406. Ministry of Health and Sports, Nay Pyi Taw, Myanmar

Maung Maung Than Htike

407. Santé publique France, Saint-Maurice, France

Laetitia Huiart

408. VU University Medical Center, Amsterdam, The Netherlands

Martijn Huisman & Natasja M. Van Schoor

409. American University of Beirut, Beirut, Lebanon

Nahla Hwalla & Abla M. Sibai

410. Cairo University, Cairo, Egypt

Mohsen M. Ibrahim

411. Erasmus Medical Center Rotterdam, Rotterdam, The Netherlands

M. Arfan Ikram & Maryam Kavousi

412. University of Valencia, Valencia, Spain

Carmen Iñiguez & Josep Redon

413. Medical University Varna, Varna, Bulgaria

Violeta Iotova

414. The University of Tokyo, Tokyo, Japan

Takafumi Ishida

415. Alex Ekwueme Federal University Teaching Hospital, Abakaliki, Nigeria

Godsent C. Isiguzo

416. The Hospital for Sick Children, Toronto, Ontario, Canada

Muhammad Islam & Zulfiqar A. Bhutta

417. Deakin University, Geelong, Victoria, Australia

Sheikh Mohammed Shariful Islam

418. Emory University, Atlanta, GA, USA

Duygu Islek, Reynaldo Martorell, Thirunavukkarasu Sathish & Aryeh D. Stein

419. Bulgarian Academy of Sciences, Sofia, Bulgaria

Ivaila Y. Ivanova-Pandourska & Yanitsa V. Zhecheva

420. Tokyo Metropolitan Institute of Gerontology, Tokyo, Japan

Masanori Iwasaki

421. Hadassah University Medical Center, Jerusalem, Israel

Jeremy M. Jacobs & Jochanan Stessman

422. Université Catholique de Louvain, Brussels, Belgium

Michel Jadoul

423. Gambia National Nutrition Agency, Banjul, The Gambia

Bakary Jallow & Modou Cheyassin Phall

424. Kuwait Institute for Scientific Research, Safat, Kuwait

Kazi M. Jamil

425. Public Health Agency of Sweden, Solna, Sweden

Anna Jansson & Kenisha Russell Jonsson

426. Norwegian University of Science and Technology, Trondheim, Norway

Imre Janszky, Steinar Krokstad & Elin P. Sørgjerd

427. University of Melbourne, Melbourne, Victoria, Australia

Edward Janus

428. Sports University of Tirana, Tirana, Albania

Juel Jarani

429. Heart Foundation, Melbourne, Victoria, Australia

Garry Jennings & Bill Stavreski

430. Guangzhou 12th Hospital, Guangzhou, China

Chao Qiang Jiang

431. Universidad Eugenio Maria de Hostos, Santo Domingo, Dominican Republic

Ramon O. Jimenez

432. Simon Fraser University, Burnaby, British Columbia, Canada

Michel Joffres

433. Institute of Molecular and Clinical Ophthalmology Basel, Basel, Switzerland

Jost B. Jonas

434. University of New South Wales, Sydney, New South Wales, Australia

Jitendra Jonnagaddala, Chandini Nekkantti, Aletta E. Schutte, Leon A. Simons & Mark Woodward

435. World Health Organization Country Office, Delhi, India

Pradeep Joshi

436. Guilan University of Medical Sciences, Rasht, Iran

Farahnaz Joukar & Fariborz Mansour-Ghanaei

437. University of Opole, Opole, Poland

Jacek J. Jóźwiak

438. Gulu University, Gulu, Uganda

Felix O. Kaducu

439. University of Crete, Heraklion, Greece

Anthony Kafatos

440. Hungarian School Sport Federation, Budapest, Hungary

Mónika Kaj & Viktoria Anna Kovacs

441. National Center for Disease Control and Public Health, Tbilisi, Georgia

Natia Kakutia, Lela Shengelia & Lela Sturua

442. Ministry of Health, Bratislava, Slovakia

Daniela Kállayová

443. Sri Venkateswara University, Tirupati, India

Kodanda R. Kanala

444. Sree Chitra Tirunal Institute for Medical Sciences and Technology, Trivandrum, India

Srinivasan Kannan

445. Hellenic Medical Association for Obesity, Athens, Greece

Efthymios Kapantais & Themistoklis Tzotzas

446. Harokopio University, Athens, Greece

Eva Karaglani, Yannis Manios & Christina Mavrogianni

447. National and Kapodistrian University of Athens, Athens, Greece

Argyro Karakosta, George S. Stergiou, Konstantinos D. Tambalis & Giota Touloumi

448. Maharajgunj Medical Campus, Kathmandu, Nepal

Khem B. Karki

449. Université Officielle de Bukavu, Bukavu, Democratic Republic of the Congo

Philippe B. Katchunga

450. Aarhus University, Aarhus, Denmark

Marzieh Katibeh, Kim Overvad & Jakob Tarp

451. Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

Joanne Katz & Moyses Szklo

452. Pennington Biomedical Research Center, Baton Rouge, LA, USA

Peter T. Katzmarzyk

453. National Institute of Epidemiology, Chennai, India

Prabhdeep Kaur & Sudha Ramachandra Rao

454. University of Toronto, Toronto, Ontario, Canada

Calvin Ke

455. University of Münster, Münster, Germany

Ulrich Keil

456. Israel Center for Disease Control, Ramat Gan, Israel

Lital Keinan Boker & Tal Shimony

457. Research Institute for Primordial Prevention of Non-communicable Disease, Isfahan, Iran

Roya Kelishadi

458. Kyrgyz State Medical Academy, Bishkek, Kyrgyzstan

Alina Kerimkulova & Erkin Mirrakhimov

459. Research Institute of Child Nutrition, Dortmund, Germany

Mathilde Kersting

460. Shahrekord University of Medical Sciences, Shahrekord, Iran

Arsalan Khaledifar

461. University of Cambridge, Cambridge, UK

Kay-Tee Khaw & Nicholas Wareham

462. Mazandaran University of Medical Sciences, Sari, Iran

Motahareh Kheradmand & Mahmood Moosazadeh

463. Hypertension Research Center, Isfahan, Iran

Alireza Khosravi

464. Medical University of Innsbruck, Innsbruck, Austria

Ursula Kiechl-Kohlendorfer, Stefan Kiechl, Michael Knoflach, Raimund Pechlaner, Hanno Ulmer, Johann Willeit & Peter Willeit

465. VASCage, Innsbruck, Austria

Sophia J. Kiechl & Stefan Kiechl

466. Muhimbili University of Health and Allied Sciences, Dar es Salaam, Tanzania

Japhet Killewo

467. Yonsei University College of Medicine, Seoul, Republic of Korea

Hyeon Chang Kim

468. National Cancer Center, Goyang-si, Republic of Korea

Jeongseon Kim & Jeonghee Lee

469. Sahlgrenska University Hospital, Gothenburg, Sweden

Jenny M. Kindblom, Claes Ohlsson & Annika Rosengren

470. Newcastle University, Newcastle, UK

Andrew Kingston & Louise Robinson

471. University College South Denmark, Haderslev, Denmark

Heidi Klakk

472. Statistics Austria, Vienna, Austria

Jeannette Klimont

473. B. P. Koirala Institute of Health Sciences, Dharan, Nepal

Bhawesh Koirala & Sanjib K. Sharma

474. University of Vienna, Vienna, Austria

Jürgen König & Petra Rust

475. Tartu University Clinics, Tartu, Estonia

Paul Korrovits & Margus Punab

476. Kansai Medical University, Hirakata, Japan

Katsuyasu Kouda, Harunobu Nakamura & Kumiko Ohara

477. District Department of State Public Health Service, Hildburghausen, Germany

Éva Kovács

478. Pontificia Universidad Católica Argentina, Buenos Aires, Argentina

Irina Kovalskys

479. Ministry of Health and Wellness, Port Louis, Mauritius

Sudhir Kowlessur

480. University Hospital Ulm, Ulm, Germany

Wolfgang Kratzer

481. Croatian Institute of Public Health, Zagreb, Croatia

Helena Krizan, Sanja Musić Milanović & Ranko Stevanović

482. Institute of Nutrition of Central America and Panama, Guatemala City, Guatemala

Maria F. Kroker-Lobos

483. North-West University, Potchefstroom, South Africa

Herculina S. Kruger & Ruan Kruger

484. South African Medical Research Council, Potchefstroom, South Africa

Herculina S. Kruger & Ruan Kruger

485. University of Physical Education, Kraków, Poland

Łukasz Kryst

486. University of Jyväskylä, Jyväskylä, Finland

Urho M. Kujala

487. Institute of Public Health, Podgorica, Montenegro

Enisa Kujundzic & Boban Mugoša

488. Amrita Institute of Medical Sciences, Cochin, India

R. Krishna Kumar & Manu Raj

489. Institute of Endocrinology, Prague, Czech Republic

Marie Kunešová, Radka Taxová Braunerová & Hana Zamrazilová

490. All India Institute of Medical Sciences, New Delhi, India

Yadlapalli S. Kusuma

491. African Population and Health Research Center, Nairobi, Kenya

Catherine Kyobutungi, Shukri F. Mohamed & Elvis O. A. Wambiya

492. Hanoi University of Public Health, Hanoi, Vietnam

Quang Ngoc La, Khuong Quynh Long, Thi Tuyet-Hanh Tran & Hoang Van Minh

493. Hassan First University of Settat, Settat, Morocco

Fatima Zahra Laamiri

494. Ministry of Health, Algiers, Algeria

Youcef Laid

495. Ministry of Health, Georgetown, Guyana

Lachmie Lall

496. Sahlgrenska Academy, Gothenburg, Sweden

Georg Lappas

497. Endocrinology and Metabolism Research Center, Tehran, Iran

Bagher Larijani

498. Clinical Research Education, Networking & Consultancy, Douala, Cameroon

Simo Pone Larissa

499. University of Public Health, Yangon, Myanmar

Tint Swe Latt & Ko Ko Zaw

500. Centro Studi Epidemiologici di Gubbio, Gubbio, Italy

Martino Laurenzi

501. National University Health System, Singapore, Singapore

Jeannette Lee, Charlie Lim, Wei-Yen Lim, Ei Ei K. Nang, Xueling Sim & E. Shyong Tai

502. University of Leicester, Leicester, UK

Paul H. Lee

503. Tampere University Hospital, Tampere, Finland

Terho Lehtimäki

504. Tampere University, Tampere, Finland

Terho Lehtimäki

505. University of Douala, Douala, Cameroon

Daniel Lemogoum

506. University of Cape Town, Cape Town, South Africa

Naomi S. Levitt

507. West Virginia University, Morgantown, WV, USA

Christa L. Lilly & William A. Neal

508. Oswaldo Cruz Foundation Rene Rachou Research Institute, Belo Horizonte, Brazil

M. Fernanda Lima-Costa & Sergio Viana Peixoto

509. Shanghai Institute of Nutrition and Health of Chinese Academy of Sciences, Shanghai, China

Xu Lin & Liang Sun

510. Uppsala University, Uppsala, Sweden

Yi-Ting Lin, Lars Lind, Johan Sundström & Agneta Yngve

511. Capital Medical University Beijing An Zhen Hospital, Beijing, China

Jing Liu & Dong Zhao

512. Taipei Medical University, Taipei, Taiwan

Wei-Cheng Lo

513. Servicio Andaluz de Salud, Sevilla, Spain

Guadalupe Longo Abril

514. Sports Medical Center of Minho, Braga, Portugal

Oscar Lopes

515. Universidad San Martín de Porres, Lima, Peru

Tania Lopez & Luis Revilla

516. Consejería de Sanidad Junta de Castilla y León, Valladolid, Spain

José Eugenio Lozano & Tomas Vega

517. Universidade Federal de Ciências da Saúde de Porto Alegre, Porto Alegre, Brazil

Janice L. Lukrafka

518. Ilembula Lutheran Hospital, Ilembula, Tanzania

Charles Lunogelo

519. University of Kinshasa Hospital, Kinshasa, Democratic Republic of the Congo

Jean-René M'Buyamba-Kabangu

520. Coimbra University Hospital Center, Coimbra, Portugal

Luisa M. Macieira & Lèlita C. Santos

521. University of Texas Rio Grande Valley, Harlingen, TX, USA

Gladys E. Maestre

522. Institute of Neuroscience of the National Research Council, Padua, Italy

Stefania Maggi & Marianna Noale

523. Baker Heart and Diabetes Institute, Melbourne, Victoria, Australia

Dianna J. Magliano & Jonathan E. Shaw

524. Agricultural University of Athens, Athens, Greece

Emmanuella Magriplis & Antonis Zampelas

525. Academia VBHC, São Paulo, Brazil

Marcia Makdisse

526. SB RAS Federal Research Center Institute of Cytology and Genetics, Novosibirsk, Russia

Sofia K. Malyutina & Yury P. Nikitin

527. Université Catholique de Bukavu, Bukavu, Democratic Republic of the Congo

Masimango Imani Manix

528. University of Northern British Columbia, Prince George, British Columbia, Canada

Taru Manyanga

529. University of Padua, Padua, Italy

Enzo Manzato & Sabina Zambon

530. University Medicine Greifswald, Greifswald, Germany

Joany Mariño, Matthias Nauck, Astrid Petersmann, Adrian Richter, Sabine Schipf, Carsten O. Schmidt & Henry Völzke

531. Hellenic Mediterranean University, Siteia, Greece

Anastasia Markaki

532. Loughborough University, Loughborough, UK

Oonagh Markey

533. Ministry of Health, Nicosia, Cyprus

Eliza Markidou Ioannidou

534. Lausanne University Hospital, Lausanne, Switzerland

Pedro Marques-Vidal, Jardena J. Puder & Peter Vollenweider

535. University of Lausanne, Lausanne, Switzerland

Pedro Marques-Vidal & Peter Vollenweider

536. Secretaria de Estado da Saúde de Santa Catarina, Florianópolis, Brazil

Larissa Pruner Marques

537. CIBERCV, Barcelona, Spain

Jaume Marrugat

538. Institut Hospital del Mar d'Investigacions Mèdiques, Barcelona, Spain

Jaume Marrugat & Miquel Porta

539. Mary Immaculate College, Limerick, Ireland

Rosemarie Martin

540. Hungarian Society of Sports Medicine, Budapest, Hungary

Eva Martos

541. Paracelsus Medical University, Salzburg, Austria

Katharina Maruszczak & Daniel Weghuber

542. Institute for Cancer Research, Prevention and Clinical Network, Florence, Italy

Giovanna Masala & Calogero Saleva

543. Universidade Estadual do Centro-Oeste, Guarapuava, Brazil

Luis P. Mascarenhas

544. Sher-i-Kashmir Institute of Medical Sciences, Srinagar, India

Shariq R. Masoodi

545. UiT The Arctic University of Norway, Tromsø, Norway

Ellisiv B. Mathiesen & Tom Wilsgaard

546. ICMR–National Centre for Disease Informatics and Research, Bengaluru, India

Prashant Mathur

547. Sefako Makgatho Health Sciences University, Pretoria, South Africa

Tandi E. Matsha

548. Centro de Estudos do Laboratório de Aptidão Física de São Caetano do Sul, São Paulo, Brazil

Victor Matsudo

549. Brown University, Providence, RI, USA

Stephen T. McGarvey

550. London School of Hygiene & Tropical Medicine, London, UK

Martin McKee, Moffat Nyirenda, Alison J. Price & Liam Smeeth

551. University of Edinburgh, Edinburgh, UK

Stela McLachlan & Jacqueline F. Price

552. Weill Cornell Medicine, New York City, NY, USA

Margaret L. McNairy

553. Institut National de la Santé et de la Recherche Médicale, Lille, France

Aline Meirhaeghe

554. Arabkir Medical Centre–Institute of Child and Adolescent Health, Yerevan, Armenia

Marina Melkumova & Yeva Movsesyan

555. Universidad de los Andes, Bogotá, Colombia

Carlos O. Mendivil & Pedro J. Quiroga-Padilla

556. Robert Koch Institute, Berlin, Germany

Gert B. M. Mensink, Angelika Schaffrath Rosario & Anja Schienkiewitz

557. University of Abidjan, Abidjan, Côte d'Ivoire

Diane T. Meto

558. Pirogov Russian National Research Medical University, Moscow, Russia

Olga Milushkina & Natalia A. Skoblina

559. Universidade de Lisboa, Lisbon, Portugal

Cláudia S. Minderico, Diana A. Santos & Luis B. Sardinha

560. Saveetha Dental Colleges & Hospitals, Chennai, India

G. K. Mini

561. Democritus University, Alexandroupolis, Greece

Daphne Mirkopoulou

562. Grigore T Popa University of Medicine and Pharmacy, Iasi, Romania

Veronica Mocanu

563. Università degli Studi di Firenze, Florence, Italy

Pietro A. Modesti

564. Ain Shams University, Cairo, Egypt

Mostafa K. Mohamed & Aya Mostafa

565. Isfahan Cardiovascular Research Center, Isfahan, Iran

Noushin Mohammadifard & Nizal Sarrafzadegan

566. Strasbourg University Hospital, Strasbourg, France

Marie Moitry

567. Mulago Hospital, Kampala, Uganda

Charles K. Mondo

568. Instituto Conmemorativo Gorgas de Estudios de la Salud, Panama City, Panama

Roger A. Montenegro Mendoza, Jorge Motta, Hedley K. Quintana & Julio Zuñiga Cisneros

569. University of Limpopo, Sovenga, South Africa

Kotsedi Daniel K. Monyeki

570. University of Medical Sciences of Cienfuegos, Cienfuegos, Cuba

Alain Morejon

571. Ministry of Health and Wellness, Belmopan, Belize

Francis Morey

572. Royal College of Surgeons in Ireland, Dublin, Ireland

Karen Morgan

573. La Trobe University, Melbourne, Victoria, Australia

George Moschonis

574. Sabzevar University of Medical Sciences, Sabzevar, Iran

Alireza Moslem & Mohammad Hossien Saghi

575. International Institute of Molecular and Cell Biology, Warsaw, Poland

Malgorzata Mossakowska & Przemysław Slusarczyk

576. World Health Organization Country Office, Lilongwe, Malawi

Kelias P. Msyamboza

577. Department of Public Health, Nay Pyi Taw, Myanmar

Thet Thet Mu

578. Albanian Sports Science Association, Tirana, Albania

Florian Muca

579. University of Brescia, Brescia, Italy

Maria L. Muiesan & Massimo Salvetti

580. University of Limerick, Limerick, Ireland

Elaine M. Murtagh

581. Makerere University School of Public Health, Kampala, Uganda

Geofrey Musinguzi & Fred Nuwaha

582. University de Kinshasa, Kinshasa, Democratic Republic of the Congo

Muel Telo M. C. Muyer

583. Bushehr University of Medical Sciences, Bushehr, Iran

Iraj Nabipour

584. Ulm University, Ulm, Germany

Gabriele Nagel

585. Suraj Eye Institute, Nagpur, India

Vinay B. Nangia

586. UNICEF, Yaoundé, Cameroon

Martin Nankap

587. Ministry of Health, Apia, Samoa

Take Naseri

588. Karolinska Institutet, Stockholm, Sweden

Martin Neovius, Michael Sjöström, Per Tynelius & Peter Ueda

589. National Institute of Hygiene and Epidemiology, Hanoi, Vietnam

Chung T. Nguyen

590. University of Medicine and Pharmacy, Ho Chi Minh City, Vietnam

Nguyen D. Nguyen & Oanh T. H. Trinh

591. Hanoi Medical University, Hanoi, Vietnam

Quang Ngoc Nguyen

592. Xi’an Jiaotong University, Xi’an, China

Peng Nie

593. LifeDoc Health, Memphis, TN, USA

Ramfis E. Nieto-Martínez

594. Heartfile, Islamabad, Pakistan

Sania Nishtar

595. Eastern Mediterranean Public Health Network, Amman, Jordan

Mohannad Al Nsour

596. University of Manchester, Manchester, UK

Terence W. O'Neill & Frederick C. Wu

597. State University of Medicine and Pharmacy, Chisinau, Moldova

Galina Obreja

598. Tachikawa General Hospital, Nagaoka, Japan

Eiji Oda

599. University of Abuja College of Health Sciences, Abuja, Nigeria

Augustine N. Odili

600. Korea Centers for Disease Control and Prevention, Cheongju-si, Republic of Korea

Kyungwon Oh & Suyeon Park

601. Japan Wildlife Research Center, Tokyo, Japan

Ryutaro Ohtsuka

602. Gadarif University, Gadarif, Sudan

Saeed M. Omar

603. Istanbul University, Istanbul, Turkey

Altan Onat

604. Ministry of Health, Bandar Seri Begawan, Brunei

Sok King Ong & Khairil Si-Ramlee

605. University of Madeira, Funchal, Portugal

Rui Ornelas

606. University of Puerto Rico, San Juan, Puerto Rico

Ana P. Ortiz & Cynthia M. Pérez

607. Osteoporosis Research Center, Tehran, Iran

Afshin Ostovar

608. Universidad de Santander, Bucaramanga, Colombia

Johanna A. Otero

609. Kwame Nkrumah University of Science and Technology, Kumasi, Ghana

Ellis Owusu-Dabo

610. University of Wisconsin-Madison, Madison, WI, USA

Alberto Palloni

611. Privatpraxis Prof Jonas und Dr Panda-Jonas, Heidelberg, Germany

Songhomitra Panda-Jonas

612. IRCCS Ente Ospedaliero Specializzato in Gastroenterologia S. de Bellis, Bari, Italy

Francesco Panza

613. Zayed University, Abu Dhabi, United Arab Emirates

Dimitrios Papandreou

614. Catholic University of Daegu, Daegu, Republic of Korea

Soon-Woo Park

615. University of Medicine, Pharmacy, Science and Technology of Târgu Mures, Târgu Mures, Romania

Ionela M. Pascanu & Raluca M. Pop

616. Jivandeep Hospital, Anand, India

Nikhil D. Patel

617. Centro de Investigação em Saúde de Angola, Caxito, Angola

João M. Pedro

618. South African Medical Research Council, Durban, South Africa

Nasheeta Peer

619. National Dental Care Centre Singapore, Singapore, Singapore

Marco A. Peres

620. University Hospital of Varese, Varese, Italy

Emanuela Pettenuzzo

621. Vietnam National Heart Institute, Hanoi, Vietnam

Son Thai Pham

622. Clínica de Medicina Avanzada Dr. Abel González, Santo Domingo, Dominican Republic

Rafael N. Pichardo

623. University of Sarajevo, Sarajevo, Bosnia and Herzegovina

Aida Pilav

624. Cardiovascular Prevention Centre Udine, Udine, Italy

Lorenza Pilotto & Diego Vanuzzo

625. University of Pisa, Pisa, Italy

Francesco Pistelli

626. Ministry of Health and Medical Services, Honiara, Solomon Islands

Freda Pitakaka

627. Public Health Agency of Catalonia, Barcelona, Spain

Pedro Plans-Rubió

628. O. M. Marzeyev Institute for Public Health of the National Academy of the Medical Sciences of Ukraine, Kyiv, Ukraine

Alina G. Platonova & Nadija S. Polka

629. Universiti Kebangsaan Malaysia, Kuala Lumpur, Malaysia

Bee Koon Poh & Jyh Eiin Wong

630. Ardabil University of Medical Sciences, Ardabil, Iran

Farhad Pourfarzi

631. Universidade Pedagógica, Maputo, Mozambique

Antonio Prista

632. Centre for Disease Prevention and Control, Riga, Latvia

Iveta Pudule & Biruta Velika

633. Sulaimani Polytechnic University, Sulaymaniyah, Iraq

Muhammed S. Qadir

634. Alborz University of Medical Sciences, Karaj, Iran

Mostafa Qorbani

635. Ministry of Health, Hanoi, Vietnam

Tran Quoc Bao

636. Pure Earth, Dhaka, Bangladesh

Mahfuzar Rahman

637. Institute of Epidemiology Disease Control and Research, Dhaka, Bangladesh

Mahmudur Rahman

638. University of Turku, Turku, Finland

Olli Raitakari

639. UNICEF, Baku, Azerbaijan

Tamerlan Rajabov

640. World Health Organization Country Office, Juba, South Sudan

Otim P. C. Ramadan

641. Instituto Federal Riograndense, Rio Grande, Brazil

Virgílio V. Ramires

642. Institut Universitari d'Investigació en Atenció Primària Jordi Gol, Girona, Spain

Rafel Ramos

643. Universiti Putra Malaysia, Serdang, Malaysia

Lekhraj Rampal

644. University of Malaya, Kuala Lumpur, Malaysia

Sanjay Rampal

645. Sotiria Hospital, Athens, Greece

Vayia Rarra

646. University of the Philippines, Manila, The Philippines

Paul Ferdinand M. Reganit & Rody G. Sy

647. Slovak Academy of Sciences, Bratislava, Slovakia

Valéria Regecová

648. University of Santa Cruz do Sul, Santa Cruz do Sul, Brazil

Jane D. P. Renner & Cézane P. Reuter

649. Nutrition Research Foundation, Barcelona, Spain

Lourdes Ribas-Barba

650. Minas Gerais State Secretariat for Health, Belo Horizonte, Brazil

Robespierre Ribeiro

651. CS S. Agustín Ibsalut, Palma, Spain

Fernando Rigo

652. Amsterdam Institute for Global Health and Development, Amsterdam, The Netherlands

Tobias F. Rinke de Wit

653. Universidade Nove de Julho, São Paulo, Brazil

Raphael M. Ritti-Dias

654. Ministerio de Salud, Panama City, Panama

Reina G. Roa

655. Public Health Agency of Canada, Ottawa, Ontario, Canada

Cynthia Robitaille & Ming-Dong Wang

656. Universidad Industrial de Santander, Bucaramanga, Colombia

Laura A. Rodríguez-Villamizar

657. Ministry of Health and Social Protection, Bogotá, Colombia

Andrea Y. Rodríguez

658. Wuqu' Kawoq, Tecpan, Guatemala

Peter Rohloff

659. GroundWork, Fläsch, Switzerland

Fabian Rohner

660. Associazione Calabrese di Epatologia, Reggio Calabria, Italy

Elisabetta L. Romeo

661. University of Minho, Braga, Portugal

Rafaela V. Rosario

662. Fiji National University, Suva, Fiji

Ian Rouse

663. GHESKIO Clinics, Port-au-Prince, Haiti

Vanessa Rouzier

664. Universidad de San Carlos, Quetzaltenango, Guatemala

Maira H. Ruano

665. National Center for Epidemiology CIBERESP, Madrid, Spain

Emma Ruiz Moreno

666. Institute of Food Sciences of the National Research Council, Avellino, Italy

Paola Russo & Alfonso Siani

667. Medical University of Gdansk, Gdansk, Poland

Marcin Rutkowski & Tomasz Zdrojewski

668. Sitaram Bhartia Institute of Science and Research, New Delhi, India

Harshpal S. Sachdev

669. Kindergarten of Avlonari, Evia, Greece

Margarita Samoutian

670. National Institute of Health, Lima, Peru

Jose Sánchez-Abanto & Carolina B. Tarqui-Mamani

671. Ministry of Health, Jakarta, Indonesia

Sandjaja

672. Catalan Department of Health, Barcelona, Spain

Susana Sans

673. Biodonostia Health Research Institute, San Sebastián, Spain

Loreto Santa Marina

674. Instituto de Saúde Ambiental, Lisbon, Portugal

Osvaldo Santos

675. Federal University of Alagoas, Maceió, Brazil

Tamara R. Santos

676. South Karelia Social and Health Care District, Lappeenranta, Finland

Jouko L. Saramies

677. National Cancer Center, Tokyo, Japan

Norie Sawada & Shoichiro Tsugane

678. University of São Paulo Clinics Hospital, São Paulo, Brazil

Marcia Scazufca

679. Hospital Italiano de Buenos Aires, Buenos Aires, Argentina

Herman Schargrodsky

680. Medical University of Vienna, Vienna, Austria

Karin Schindler, Thomas Waldhör & Kurt Widhalm

681. Rigshospitalet, Copenhagen, Denmark

Ida Maria Schmidt

682. Academic Medical Center of University of Amsterdam, Amsterdam, The Netherlands

Constance Schultsz

683. German Institute of Human Nutrition Potsdam-Rehbruecke, Nuthetal, Germany

Matthias B. Schulze

684. The George Institute for Global Health, Sydney, New South Wales, Australia

Aletta E. Schutte

685. Center for Oral Health Services and Research Mid-Norway, Trondheim, Norway

Abhijit Sen

686. Lagos State University College of Medicine, Lagos, Nigeria

Idowu O. Senbanjo

687. University of Las Palmas de Gran Canaria, Las Palmas de Gran Canaria, Spain

Luis Serra-Majem

688. Comenius University, Bratislava, Slovakia

Ľudmila Ševčíková & Ľubica Tichá

689. Teikyo University, Tokyo, Japan

Hana Shimizu-Furusawa

690. Finnish Institute of Occupational Health, Helsinki, Finland

Rahman Shiri

691. Rutgers University, New Brunswick, NJ, USA

Labros S. Sidossis

692. National Agency for Public Health, Chisinau, Moldova

Natalia Silitrari

693. St Vincent's Hospital, Sydney, New South Wales, Australia

Judith Simons

694. Nes Municipality, Årnes, Norway

Gry Skodje

695. Health Polytechnic Jakarta II Institute, Jakarta, Indonesia

Moesijanti Y. E. Soekatri

696. Diponegoro University, Semarang, Indonesia

Agustinus Soemantri

697. University of Bari, Bari, Italy

Vincenzo Solfrizzi

698. Institut Régional de Santé Publique, Ouidah, Benin

Charles Sossa Jérome

699. University of Bordeaux, Bordeaux, France

Aïcha Soumaré & Christophe Tzourio

700. University of Hohenheim, Stuttgart, Germany

Alfonso Sousa-Poza

701. Oslo Metropolitan University, Oslo, Norway

Bente Sparboe-Nilsen & Liv Elin Torheim

702. Institute of Public Health, Skopje, North Macedonia

Igor Spiroski

703. Ss. Cyril and Methodius University, Skopje, North Macedonia

Igor Spiroski

704. Lamprecht und Stamm Sozialforschung und Beratung AG, Zurich, Switzerland

Hanspeter Stamm

705. Bonn University, Bonn, Germany

Peter Stehle

706. National Institute of Public Health–National Institute of Hygiene, Warsaw, Poland

Jakub Stokwiszewski, Lucjan Szponar & Bogdan Wojtyniak

707. Kalina Malina Kindergarten, Pazardjik, Bulgaria

Ekaterina Stoyanova

708. The Jikei University School of Medicine, Tokyo, Japan

Machi Suka

709. Fu Jen Catholic University, Taipei, Taiwan

Chien-An Sun & San-Lin You

710. University of Jordan, Amman, Jordan

Nabil William G. Sweis & Ayman A. Zayed

711. National Statistical Office, Praia, Cabo Verde

René Charles Sylva

712. Monash University, Melbourne, Victoria, Australia

Eng Joo Tan

713. Scientific Research Institute of Maternal and Child Health, Ashgabat, Turkmenistan

Maya Tanrygulyyeva

714. University of Lincoln, Lincoln, UK

Frank Tanser

715. Ministry of Health, Amman, Jordan

Mohammed Rasoul Tarawneh

716. UNICEF, Niamey, Niger

Félicité Tchibindat

717. University of Applied Sciences Utrecht, Utrecht, The Netherlands

Saskia Te Velde

718. University Medical Center Utrecht, Utrecht, The Netherlands

Erik J. Timmermans

719. National Research and Innovation Agency, Jakarta, Indonesia

Dwi H. Tjandrarini

720. Health Service, Murcia, Spain

María José Tormo

721. Institut d’Investigacio Sanitaria Illes Balears, Menorca, Spain

Maties Torrent

722. University of Bologna, Bologna, Italy

Stefania Toselli

723. Children’s Hospital of Eastern Ontario Research Institute, Ottawa, Ontario, Canada

Mark S. Tremblay

724. Hellenic Health Foundation, Athens, Greece

Antonia Trichopoulou

725. Government Medical College, Bhavnagar, India

Atul Trivedi

726. Institute of Epidemiology and Preventive Medicine, Taipei, Taiwan

Yu-Hsiang Tsao

727. Sefako Makgatho Health Sciences University, Ga-Rankuwa, South Africa

Lechaba Tshepo

728. Department of Health, Faga’alu, American Samoa

John Tuitele

729. LBJ Hospital, Faga’alu, American Samoa

John Tuitele

730. Addis Ababa University, Addis Ababa, Ethiopia

Fikru Tullu

731. Ministry of Health, Wellington, New Zealand

Maria L. Turley

732. Israel Defense Forces Medical Corps, Tel HaShomer, Israel

Gilad Twig

733. Universidad Centro–Occidental Lisandro Alvarado, Barquisimeto, Venezuela

Eunice Ugel

734. Meharry Medical College, Nashville, TN, USA

Flora A. M. Ukoli

735. University of Tampere Tays Eye Center, Tampere, Finland

Hannu M. T. Uusitalo

736. Sabiha Gokcen Ilkokulu, Ankara, Turkey

Nalan Uysal

737. Polytechnic Institute of Porto, Porto, Portugal

Susana Vale

738. Icahn School of Medicine at Mount Sinai, New York City, NY, USA

Damaskini Valvi

739. George Washington University, Washington, DC, USA

Rob M. van Dam

740. Universidad CEU San Pablo, Madrid, Spain

Gregorio Varela-Moreiras

741. Institute of Clinical Physiology of National Research Council, Pisa, Italy

Giovanni Viegi

742. Universidad San Francisco de Quito, Quito, Ecuador

Monica Vilar & Martha Yépez García

743. University Miguel Hernandez, Alicante, Spain

Jesus Vioque

744. Université de Lorraine, Nancy, France

Sophie Visvikis-Siest

745. Sunflower Nursery School, Craiova, Romania

Mihaela Vladulescu

746. North Karelia Center for Public Health, Joensuu, Finland

Tiina Vlasoff

747. University of the Witwatersrand, Johannesburg, South Africa

Alisha N. Wade

748. Institute for Medical Research, Kuala Lumpur, Malaysia

Wan Nazaimoon Wan Mohamud

749. Xinjiang Medical University, Urumqi, China

Qian Wang

750. Shanghai Educational Development Co. Ltd, Shanghai, China

Xiangjun Wang & Yang Yang

751. Ministry of Health and Welfare, Taipei, Taiwan

Ying-Wei Wang

752. Ministry of Health and Wellness, Kingston, Jamaica

Karen Webster-Kerr

753. Örebro University, Örebro, Sweden

Bo Werner

754. St George's, University of London, London, UK

Peter H. Whincup

755. Universitas Indonesia, Jakarta, Indonesia

Indah S. Widyahening

756. Rehamed-Center, Tajęcina, Poland

Rusek Wojciech

757. National Yang Ming Chiao Tung University, Taipei, Taiwan

Hon-Yen Wu

758. Institute of Food and Nutrition Development of Ministry of Agriculture and Rural Affairs, Beijing, China

Haiquan Xu

759. Beijing Institute of Ophthalmology, Beijing, China

Liang Xu

760. Children's Hospital of Fudan University, Shanghai, China

Weili Yan

761. University of Cyprus, Nicosia, Cyprus

Panayiotis K. Yiallouros

762. Niigata University, Niigata, Japan

Akihiro Yoshihara

763. South China Institute of Environmental Sciences, Guangzhou, China

Yunjiang Yu

764. International Medical University, Shah Alam, Malaysia

Safiah Md Yusof

765. Hellenic Mediterranean University, Heraklion, Greece

Vassilis Zafiropulos

766. Iran University of Medical Sciences, Tehran, Iran

Farhad Zamani

767. Center for Diabetes and Endocrine Care, Srinagar, India

Abdul Hamid Zargar

768. Jagiellonian University, Kraków, Poland

Magdalena Żegleń

769. Duke University, Durham, NC, USA

Yi Zeng

770. Peking University First Hospital, Beijing, China

Luxia Zhang & Ming-Hui Zhao

771. Jiangsu Provincial Center for Disease Control and Prevention, Nanjing, China

Shiqi Zhen

772. West Kazakhstan Medical University, Aktobe, Kazakhstan

Bekbolat Zholdin

773. Inner Mongolia Medical University, Hohhot, China

Dan Zhu

774. Przedszkole No. 81, Warsaw, Poland

Monika Zuziak

775. Johns Hopkins University, Baltimore, MD, USA

Robert E. Black

Consortia

NCD Risk Factor Collaboration (NCD-RisC)

- Anu Mishra
- , Bin Zhou
- , Andrea Rodriguez-Martinez
- , Honor Bixby
- , Rosie K. Singleton
- , Rodrigo M. Carrillo-Larco
- , Kate E. Sheffer
- , Christopher J. Paciorek
- , James E. Bennett
- , Victor Lhoste
- , Maria L. C. Iurilli

- , Mariachiara Di Cesare
- , James Bentham
- , Nowell H. Phelps
- , Marisa K. Sophiea
- , Gretchen A. Stevens
- , Goodarz Danaei
- , Melanie J. Cowan
- , Stefan Savin
- , Leanne M. Riley
- , Edward W. Gregg
- , Wichai Aekplakorn
- , Noor Ani Ahmad
- , Jennifer L. Baker
- , Adela Chirita-Emandi
- , Farshad Farzadfar
- , Günther Fink
- , Mirjam Heinen
- , Nayu Ikeda
- , Andre P. Kengne
- , Young-Ho Khang
- , Tiina Laatikainen
- , Avula Laxmaiah
- , Jun Ma
- , Michele Monroy-Valle
- , Malay K. Mridha
- , Cristina P. Padez
- , Andrew Reynolds
- , Maroje Sorić
- , Gregor Starc
- , James P. Wirth
- , Leandra Abarca-Gómez
- , Ziad A. Abdeen
- , Shynar Abdrakhmanova
- , Suhaila Abdul Ghaffar
- , Hanan F. Abdul Rahim
- , Zulfiya Abdurrahmonova
- , Niveen M. Abu-Rmeileh

- , Jamila Abubakar Garba
- , Benjamin Acosta-Cazares
- , Ishag Adam
- , Marzena Adamczyk
- , Robert J. Adams
- , Seth Adu-Afarwuah
- , Kaosar Afsana
- , Shoaib Afzal
- , Valirie N. Agbor
- , Imelda A. Agdeppa
- , Javad Aghazadeh-Attari
- , Hassan Aguenaou
- , Carlos A. Aguilar-Salinas
- , Charles Agyemang
- , Mohamad Hasnan Ahmad
- , Ali Ahmadi
- , Naser Ahmadi
- , Nastaran Ahmadi
- , Imran Ahmed
- , Soheir H. Ahmed
- , Wolfgang Ahrens
- , Gulmira Aitmurzaeva
- , Kamel Ajlouni
- , Hazzaa M. Al-Hazzaa
- , Badreya Al-Lahou
- , Rajaa Al-Raddadi
- , Huda M. Al Hourani
- , Nawal M. Al Qaoud
- , Monira Alarouj
- , Fadia AlBuhairan
- , Shahla AlDhukair
- , Maryam A. Aldwairji
- , Sylvia Alexius
- , Mohamed M. Ali
- , Abdullah Alkandari
- , Ala’a Alkerwi
- , Buthaina M. Alkhatib

- , Kristine Allin
- , Mar Alvarez-Pedrerol
- , Eman Aly
- , Deepak N. Amarapurkar
- , Pilar Amiano Etxezarreta
- , John Amoah
- , Norbert Amougou
- , Philippe Amouyel
- , Lars Bo Andersen
- , Sigmund A. Anderssen
- , Odysseas Androutsos
- , Lars Ängquist
- , Ranjit Mohan Anjana
- , Alireza Ansari-Moghaddam
- , Elena Anufrieva
- , Hajer Aounallah-Skhiri
- , Joana Araújo
- , Inger Ariansen
- , Tahir Aris
- , Raphael E. Arku
- , Nimmathota Arlappa
- , Krishna K. Aryal
- , Nega Aseffa
- , Thor Aspelund
- , Felix K. Assah
- , Batyrbek Assembekov
- , Maria Cecília F. Assunção
- , May Soe Aung
- , Juha Auvinen
- , Mária Avdičová
- , Shina Avi
- , Ana Azevedo
- , Mohsen Azimi-Nezhad
- , Fereidoun Azizi
- , Mehrdad Azmin
- , Bontha V. Babu
- , Maja Bæksgaard Jørgensen

- , Azli Baharudin
- , Suhad Bahijri
- , Marta Bakacs
- , Nagalla Balakrishna
- , Yulia Balanova
- , Mohamed Bamoshmoosh
- , Maciej Banach
- , José R. Banegas
- , Joanna Baran
- , Rafał Baran
- , Carlo M. Barbagallo
- , Valter Barbosa Filho
- , Alberto Barceló
- , Maja Baretić
- , Amina Barkat
- , Joaquin Barnoya
- , Lena Barrera
- , Marta Barreto
- , Aluisio J. D. Barros
- , Mauro Virgílio Gomes Barros
- , Anna Bartosiewicz
- , Abdul Basit
- , Joao Luiz D. Bastos
- , Iqbal Bata
- , Anwar M. Batieha
- , Aline P. Batista
- , Rosangela L. Batista
- , Zhamilya Battakova
- , Louise A. Baur
- , Pascal M. Bayauli
- , Robert Beaglehole
- , Silvia Bel-Serrat
- , Antonisamy Belavendra
- , Habiba Ben Romdhane
- , Judith Benedics
- , Mikhail Benet
- , Gilda Estela Benitez Rolandi

- , Elling Bere
- , Ingunn Holden Bergh
- , Yemane Berhane
- , Salim Berkinbayev
- , Antonio Bernabe-Ortiz
- , Gailute Bernotiene
- , Ximena Berrios Carrasola
- , Heloísa Bettiol
- , Manfred E. Beutel
- , Augustin F. Beybey
- , Jorge Bezerra
- , Aroor Bhagyalaxmi
- , Sumit Bharadwaj
- , Santosh K. Bhargava
- , Hongsheng Bi
- , Yufang Bi
- , Daniel Bia
- , Katia Biasch
- , Elysée Claude Bika Lele
- , Mukharram M. Bikbov
- , Bihungum Bista
- , Dusko J. Bjelica
- , Anne A. Bjerregaard
- , Peter Bjerregaard
- , Espen Bjertness
- , Marius B. Bjertness
- , Cecilia Björkelund
- , Katia V. Bloch
- , Anneke Blokstra
- , Moran Blychfeld Magnazu
- , Simona Bo
- , Martin Bobak
- , Lynne M. Boddy
- , Bernhard O. Boehm
- , Jolanda M. A. Boer
- , Jose G. Boggia
- , Elena Bogova

- , Carlos P. Boissonnet
- , Stig E. Bojesen
- , Marialaura Bonaccio
- , Vanina Bongard
- , Alice Bonilla-Vargas
- , Matthias Bopp
- , Herman Borghs
- , Pascal Bovet
- , Khadichamo Boymatova
- , Lien Braeckevelt
- , Lutgart Braeckman
- , Marjolijn C. E. Bragt
- , Imperia Brajkovich
- , Francesco Branca
- , Juergen Breckenkamp
- , João Breda
- , Hermann Brenner
- , Lizzy M. Brewster
- , Garry R. Brian
- , Yajaira Briceño
- , Lacramioara Brinduse
- , Miguel Brito
- , Sinead Brophy
- , Johannes Brug
- , Graziella Bruno
- , Anna Bugge
- , Frank Buntinx
- , Marta Buoncristiano
- , Genc Burazeri
- , Con Burns
- , Antonio Cabrera de León
- , Joseph Cacciottolo
- , Hui Cai
- , Roberta B. Caixeta
- , Tilema Cama
- , Christine Cameron
- , José Camolas

- , Günay Can
- , Ana Paula C. Cândido
- , Felicia Cañete
- , Mario V. Capanzana
- , Naděžda Čapková
- , Eduardo Capuano
- , Rocco Capuano
- , Vincenzo Capuano
- , Marloes Cardol
- , Viviane C. Cardoso
- , Axel C. Carlsson
- , Esteban Carmuega
- , Joana Carvalho
- , José A. Casajús
- , Felipe F. Casanueva
- , Maribel Casas
- , Ertugrul Celikcan
- , Laura Censi
- , Marvin Cervantes-Loaiza
- , Juraci A. Cesar
- , Snehalatha Chamukuttan
- , Angelique Chan
- , Queenie Chan
- , Himanshu K. Chaturvedi
- , Nish Chaturvedi
- , Norsyamlina Che Abdul Rahim
- , Miao Li Chee
- , Chien-Jen Chen
- , Fangfang Chen
- , Huashuai Chen
- , Shuohua Chen
- , Zhengming Chen
- , Ching-Yu Cheng
- , Yiling J. Cheng
- , Bahman Cheraghian
- , Angela Chetrit
- , Ekaterina Chikova-Iscener

- , Mai J. M. Chinapaw
- , Anne Chinnock
- , Arnaud Chiolero
- , Shu-Ti Chiou
- , María-Dolores Chirlaque
- , Belong Cho
- , Kaare Christensen
- , Diego G. Christofaro
- , Jerzy Chudek
- , Renata Cifkova
- , Michelle Cilia
- , Eliza Cinteza
- , Massimo Cirillo
- , Frank Claessens
- , Janine Clarke
- , Els Clays
- , Emmanuel Cohen
- , Laura-María Compañ-Gabucio
- , Hans Concin
- , Susana C. Confortin
- , Cyrus Cooper
- , Tara C. Coppinger
- , Eva Corpeleijn
- , Lilia Yadira Cortés
- , Simona Costanzo
- , Dominique Cottel
- , Chris Cowell
- , Cora L. Craig
- , Amelia C. Crampin
- , Amanda J. Cross
- , Ana B. Crujeiras
- , Juan J. Cruz
- , Tamás Csányi
- , Semánová Csilla
- , Alexandra M. Cucu
- , Liufu Cui
- , Felipe V. Cureau

- , Sarah Cuschieri
- , Ewelina Czenczek-Lewandowska
- , Graziella D’Arrigo
- , Eleonora d’Orsi
- , Liliana Dacica
- , Jean Dallongeville
- , Albertino Damasceno
- , Camilla T. Damsgaard
- , Rachel Dankner
- , Thomas M. Dantoft
- , Parasmani Dasgupta
- , Saeed Dastgiri
- , Luc Dauchet
- , Kairat Davletov
- , Maria Alice Altenburg de Assis
- , Guy De Backer
- , Dirk De Bacquer
- , Amalia De Curtis
- , Patrícia de Fragas Hinnig
- , Giovanni de Gaetano
- , Stefaan De Henauw
- , Pilar De Miguel-Etayo
- , Paula Duarte de Oliveira
- , David De Ridder
- , Karin De Ridder
- , Susanne R. de Rooij
- , Delphine De Smedt
- , Mohan Deepa
- , Alexander D. Deev
- , Vincent DeGennaro Jr
- , Hélène Delisle
- , Francis Delpeuch
- , Stefaan Demarest
- , Elaine Dennison
- , Katarzyna Dereń
- , Valérie Deschamps
- , Meghnath Dhimal

- , Augusto Di Castelnuovo
- , Juvenal Soares Dias-da-Costa
- , María Elena Díaz-Sánchez
- , Alejandro Diaz
- , Pedro Díaz Fernández
- , María Pilar Díez Ripollés
- , Zivka Dika
- , Shirin Djalalinia
- , Visnja Djordjic
- , Ha T. P. Do
- , Annette J. Dobson
- , Liria Dominguez
- , Maria Benedetta Donati
- , Chiara Donfrancesco
- , Guanghui Dong
- , Yanhui Dong
- , Silvana P. Donoso
- , Angela Döring
- , Maria Dorobantu
- , Ahmad Reza Dorosty
- , Kouamelan Doua
- , Nico Dragano
- , Wojciech Drygas
- , Jia Li Duan
- , Charmaine A. Duante
- , Priscilla Duboz
- , Vesselka L. Duleva
- , Virginija Dulskiene
- , Samuel C. Dumith
- , Anar Dushpanova
- , Azhar Dyussupova
- , Vilnis Dzerve
- , Elzbieta Dziankowska-Zaborszczyk
- , Guadalupe Echeverría
- , Ricky Eddie
- , Ebrahim Eftekhar
- , Eruke E. Egbagbe

- , Robert Eggertsen
- , Sareh Eghtesad
- , Gabriele Eiben
- , Ulf Ekelund
- , Mohammad El-Khateeb
- , Laila El Ammari
- , Jalila El Ati
- , Denise Eldemire-Shearer
- , Marie Eliasen
- , Paul Elliott
- , Ronit Endevelt
- , Reina Engle-Stone
- , Rajiv T. Erasmus
- , Raimund Erbel
- , Cihangir Erem
- , Gul Ergor
- , Louise Eriksen
- , Johan G. Eriksson
- , Jorge Escobedo-de la Peña
- , Saeid Eslami
- , Ali Esmaeili
- , Alun Evans
- , David Faeh
- , Ildar Fakhradiyev
- , Albina A. Fakhretdinova
- , Caroline H. Fall
- , Elnaz Faramarzi
- , Mojtaba Farjam
- , Victoria Farrugia Sant'Angelo
- , Mohammad Reza Fattahi
- , Asher Fawwad
- , Wafaie W. Fawzi
- , Edit Feigl
- , Francisco J. Felix-Redondo
- , Trevor S. Ferguson
- , Romulo A. Fernandes
- , Daniel Fernández-Bergés

- , Daniel Ferrante
- , Thomas Ferrao
- , Gerson Ferrari
- , Marika Ferrari
- , Marco M. Ferrario
- , Catterina Ferreccio
- , Haroldo S. Ferreira
- , Eldridge Ferrer
- , Jean Ferrieres
- , Thamara Hubler Figueiró
- , Anna Fijalkowska
- , Mauro Fisberg
- , Krista Fischer
- , Leng Huat Foo
- , Maria Forsner
- , Heba M. Fouad
- , Damian K. Francis
- , Maria do Carmo Franco
- , Zlatko Fras
- , Guillermo Frontera
- , Flavio D. Fuchs
- , Sandra C. Fuchs
- , Isti I. Fujiati
- , Yuki Fujita
- , Matsuda Fumihiko
- , Viktoriya Furdela
- , Takuro Furusawa
- , Zbigniew Gaciong
- , Mihai Gafencu
- , Manuel Galán Cuesta
- , Andrzej Galbarczyk
- , Henrike Galenkamp
- , Daniela Galeone
- , Myriam Galfo
- , Fabio Galvano
- , Jingli Gao
- , Pei Gao

- , Manoli Garcia-de-la-Hera
- , María José García Mérida
- , Marta García Solano
- , Dickman Gareta
- , Sarah P. Garnett
- , Jean-Michel Gaspoz
- , Magda Gasull
- , Adroaldo Cesar Araujo Gaya
- , Anelise Reis Gaya
- , Andrea Gazzinelli
- , Ulrike Gehring
- , Harald Geiger
- , Johanna M. Geleijnse
- , Ronnie George
- , Ebrahim Ghaderi
- , Ali Ghanbari
- , Erfan Ghasemi
- , Oana-Florentina Gheorghe-Fronea
- , Alessandro Gialluisi
- , Simona Giampaoli
- , Francesco Gianfagna
- , Christian Gieger
- , Tiffany K. Gill
- , Jonathan Giovannelli
- , Glen Gironella
- , Aleksander Giwercman
- , Konstantinos Gkiouras
- , Natalya Glushkova
- , Natalja Gluškova
- , Ramesh Godara
- , Justyna Godos
- , Sibel Gogen
- , Marcel Goldberg
- , David Goltzman
- , Georgina Gómez
- , Jesús Humberto Gómez Gómez
- , Luis F. Gomez

- , Santiago F. Gómez
- , Aleksandra Gomula
- , Bruna Gonçalves Cordeiro da Silva
- , Helen Gonçalves
- , Mauer Gonçalves
- , Ana D. González-Alvarez
- , David A. Gonzalez-Chica
- , Esther M. González-Gil
- , Marcela Gonzalez-Gross
- , Margot González-Leon
- , Juan P. González-Rivas
- , Clicerio González-Villalpando
- , María-Elena González-Villalpando
- , Angel R. Gonzalez
- , Frederic Gottrand
- , Antonio Pedro Graça
- , Sidsel Graff-Iversen
- , Dušan Grafnetter
- , Aneta Grajda
- , Maria G. Grammatikopoulou
- , Ronald D. Gregor
- , Maria João Gregório
- , Else Karin Grøholt
- , Anders Grøntved
- , Giuseppe Grosso
- , Gabriella Gruden
- , Dongfeng Gu
- , Viviana Guajardo
- , Emanuela Gualdi-Russo
- , Pilar Guallar-Castillón
- , Andrea Gualtieri
- , Elias F. Gudmundsson
- , Vilmundur Gudnason
- , Ramiro Guerrero
- , Idris Guessous
- , Andre L. Guimaraes
- , Martin C. Gulliford

- , Johanna Gunnlaugsdottir
- , Marc J. Gunter
- , Xiu-Hua Guo
- , Yin Guo
- , Prakash C. Gupta
- , Rajeev Gupta
- , Oye Gureje
- , Enrique Gutiérrez González
- , Laura Gutierrez
- , Felix Gutzwiller
- , Xinyi Gwee
- , Seongjun Ha
- , Farzad Hadaegh
- , Charalambos A. Hadjigeorgiou
- , Rosa Haghshenas
- , Hamid Hakimi
- , Jytte Halkjær
- , Ian R. Hambleton
- , Behrooz Hamzeh
- , Willem A. Hanekom
- , Dominique Hange
- , Abu A. M. Hanif
- , Sari Hantunen
- , Jie Hao
- , Carla Menêses Hardman
- , Rachakulla Hari Kumar
- , Tina Harmer Lassen
- , Javad Harooni
- , Seyed Mohammad Hashemi-Shahri
- , Maria Hassapidou
- , Jun Hata
- , Teresa Haugsgjerd
- , Alison J. Hayes
- , Jiang He
- , Yuan He
- , Yuna He
- , Regina Heidinger-Felső

- , Margit Heier
- , Tatjana Hejgaard
- , Marleen Elisabeth Hendriks
- , Rafael dos Santos Henrique
- , Ana Henriques
- , Leticia Hernandez Cadena
- , Sauli Herrala
- , Marianella Herrera-Cuenca
- , Victor M. Herrera
- , Isabelle Herter-Aeberli
- , Karl-Heinz Herzig
- , Ramin Heshmat
- , Allan G. Hill
- , Sai Yin Ho
- , Suzanne C. Ho
- , Michael Hobbs
- , Doroteia A. Höfelmann
- , Michelle Holdsworth
- , Reza Homayounfar
- , Clara Homs
- , Wilma M. Hopman
- , Andrea R. V. R. Horimoto
- , Claudia M. Hormiga
- , Bernardo L. Horta
- , Leila Houti
- , Christina Howitt
- , Thein Thein Htay
- , Aung Soe Htet
- , Maung Maung Than Htike
- , Yonghua Hu
- , José María Huerta
- , Ilpo Tapani Huhtaniemi
- , Laetitia Huiart
- , Constanta Huidumac Petrescu
- , Martijn Huisman
- , Abdullatif Husseini
- , Chinh Nguyen Huu

- , Inge Huybrechts
- , Nahla Hwalla
- , Jolanda Hyska
- , Licia Iacoviello
- , Ellina M. Iakupova
- , Jesús M. Ibarluzea
- , Mohsen M. Ibrahim
- , Norazizah Ibrahim Wong
- , M. Arfan Ikram
- , Carmen Iñiguez
- , Violeta Iotova
- , Vilma E. Irazola
- , Takafumi Ishida
- , Godsent C. Isiguzo
- , Muhammad Islam
- , Sheikh Mohammed Shariful Islam
- , Duygu Islek
- , Ivaila Y. Ivanova-Pandourska
- , Masanori Iwasaki
- , Tuija Jääskeläinen
- , Rod T. Jackson
- , Jeremy M. Jacobs
- , Michel Jadoul
- , Tazeen Jafar
- , Bakary Jallow
- , Kenneth James
- , Kazi M. Jamil
- , Konrad Jamrozik
- , Anna Jansson
- , Imre Janszky
- , Edward Janus
- , Juel Jarani
- , Marjo-Riitta Jarvelin
- , Grazyna Jasienska
- , Ana Jelaković
- , Bojan Jelaković
- , Garry Jennings

- , Chao Qiang Jiang
- , Ramon O. Jimenez
- , Karl-Heinz Jöckel
- , Michel Joffres
- , Jari J. Jokelainen
- , Jost B. Jonas
- , Jitendra Jonnagaddala
- , Torben Jørgensen
- , Pradeep Joshi
- , Josipa Josipović
- , Farahnaz Joukar
- , Jacek J. Jóźwiak
- , Debra S. Judge
- , Anne Juolevi
- , Gregor Jurak
- , Iulia Jurca Simina
- , Vesna Juresa
- , Rudolf Kaaks
- , Felix O. Kaducu
- , Anthony Kafatos
- , Mónika Kaj
- , Eero O. Kajantie
- , Natia Kakutia
- , Daniela Kállayová
- , Zhanna Kalmatayeva
- , Ofra Kalter-Leibovici
- , Yves Kameli
- , Freja B. Kampmann
- , Kodanda R. Kanala
- , Srinivasan Kannan
- , Efthymios Kapantais
- , Eva Karaglani
- , Argyro Karakosta
- , Line L. Kårhus
- , Khem B. Karki
- , Philippe B. Katchunga
- , Marzieh Katibeh

- , Joanne Katz
- , Peter T. Katzmarzyk
- , Jussi Kauhanen
- , Prabhdeep Kaur
- , Maryam Kavousi
- , Gyulli M. Kazakbaeva
- , François F. Kaze
- , Calvin Ke
- , Ulrich Keil
- , Lital Keinan Boker
- , Sirkka Keinänen-Kiukaanniemi
- , Roya Kelishadi
- , Cecily Kelleher
- , Han C. G. Kemper
- , Maryam Keramati
- , Alina Kerimkulova
- , Mathilde Kersting
- , Timothy Key
- , Yousef Saleh Khader
- , Arsalan Khaledifar
- , Davood Khalili
- , Kay-Tee Khaw
- , Bahareh Kheiri
- , Motahareh Kheradmand
- , Alireza Khosravi
- , Ilse M. S. L. Khouw
- , Ursula Kiechl-Kohlendorfer
- , Sophia J. Kiechl
- , Stefan Kiechl
- , Japhet Killewo
- , Hyeon Chang Kim
- , Jeongseon Kim
- , Jenny M. Kindblom
- , Andrew Kingston
- , Heidi Klakk
- , Magdalena Klimek
- , Jeannette Klimont

- , Jurate Klumbiene
- , Michael Knoflach
- , Bhawesh Koirala
- , Elin Kolle
- , Patrick Kolsteren
- , Jürgen König
- , Raija Korpelainen
- , Paul Korrovits
- , Magdalena Korzycka
- , Jelena Kos
- , Seppo Koskinen
- , Katsuyasu Kouda
- , Éva Kovács
- , Viktoria Anna Kovacs
- , Irina Kovalskys
- , Sudhir Kowlessur
- , Slawomir Koziel
- , Jana Kratenova
- , Wolfgang Kratzer
- , Vilma Kriaucioniene
- , Susi Kriemler
- , Peter Lund Kristensen
- , Helena Krizan
- , Maria F. Kroker-Lobos
- , Steinar Krokstad
- , Daan Kromhout
- , Herculina S. Kruger
- , Ruan Kruger
- , Łukasz Kryst
- , Ruzena Kubinova
- , Renata Kuciene
- , Urho M. Kujala
- , Enisa Kujundzic
- , Zbigniew Kulaga
- , Mukhtar Kulimbet
- , R. Krishna Kumar
- , Marie Kunešová

- , Pawel Kurjata
- , Yadlapalli S. Kusuma
- , Vladimir Kutsenko
- , Kari Kuulasmaa
- , Catherine Kyobutungi
- , Quang Ngoc La
- , Fatima Zahra Laamiri
- , Carl Lachat
- , Karl J. Lackner
- , Youcef Laid
- , Lachmie Lall
- , Tai Hing Lam
- , Maritza Landaeta Jimenez
- , Edwige Landais
- , Vera Lanska
- , Georg Lappas
- , Bagher Larijani
- , Simo Pone Larissa
- , Tint Swe Latt
- , Martino Laurenzi
- , Laura Lauria
- , Maria Lazo-Porras
- , Gwenaëlle Le Coroller
- , Khanh Le Nguyen Bao
- , Agnès Le Port
- , Tuyen D. Le
- , Jeannette Lee
- , Jeonghee Lee
- , Paul H. Lee
- , Nils Lehmann
- , Terho Lehtimäki
- , Daniel Lemogoum
- , Branimir Leskošek
- , Justyna Leszczak
- , Katja B. Leth-Møller
- , Gabriel M. Leung
- , Naomi S. Levitt

- , Yanping Li
- , Merike Liivak
- , Christa L. Lilly
- , Charlie Lim
- , Wei-Yen Lim
- , M. Fernanda Lima-Costa
- , Hsien-Ho Lin
- , Xu Lin
- , Yi-Ting Lin
- , Lars Lind
- , Vijaya Lingam
- , Birgit Linkohr
- , Allan Linneberg
- , Lauren Lissner
- , Mieczyslaw Litwin
- , Jing Liu
- , Lijuan Liu
- , Wei-Cheng Lo
- , Helle-Mai Loit
- , Khuong Quynh Long
- , Guadalupe Longo Abril
- , Luis Lopes
- , Marcus V. V. Lopes
- , Oscar Lopes
- , Esther Lopez-Garcia
- , Tania Lopez
- , Paulo A. Lotufo
- , José Eugenio Lozano
- , Janice L. Lukrafka
- , Dalia Luksiene
- , Annamari Lundqvist
- , Nuno Lunet
- , Charles Lunogelo
- , Michala Lustigová
- , Edyta Łuszczki
- , Jean-René M'Buyamba-Kabangu
- , Guansheng Ma

- , Xu Ma
- , George L. L. Machado-Coelho
- , Aristides M. Machado-Rodrigues
- , Enguerran Macia
- , Luisa M. Macieira
- , Ahmed A. Madar
- , Anja L. Madsen
- , Gladys E. Maestre
- , Stefania Maggi
- , Dianna J. Magliano
- , Sara Magnacca
- , Emmanuella Magriplis
- , Gowri Mahasampath
- , Bernard Maire
- , Marjeta Majer
- , Marcia Makdisse
- , Päivi Mäki
- , Fatemeh Malekzadeh
- , Reza Malekzadeh
- , Rahul Malhotra
- , Kodavanti Mallikharjuna Rao
- , Sofia K. Malyutina
- , Lynell V. Maniego
- , Yannis Manios
- , Masimango Imani Manix
- , Jim I. Mann
- , Fariborz Mansour-Ghanaei
- , Taru Manyanga
- , Enzo Manzato
- , Anie Marcil
- , Paula Margozzini
- , Joany Mariño
- , Anastasia Markaki
- , Oonagh Markey
- , Eliza Markidou Ioannidou
- , Pedro Marques-Vidal
- , Larissa Pruner Marques

- , Jaume Marrugat
- , Yves Martin-Prevel
- , Rosemarie Martin
- , Reynaldo Martorell
- , Eva Martos
- , Katharina Maruszczak
- , Stefano Marventano
- , Giovanna Masala
- , Luis P. Mascarenhas
- , Shariq R. Masoodi
- , Ellisiv B. Mathiesen
- , Prashant Mathur
- , Alicia Matijasevich
- , Piotr Matłosz
- , Tandi E. Matsha
- , Victor Matsudo
- , Christina Mavrogianni
- , Artur Mazur
- , Jean Claude N. Mbanya
- , Shelly R. McFarlane
- , Stephen T. McGarvey
- , Martin McKee
- , Stela McLachlan
- , Rachael M. McLean
- , Scott B. McLean
- , Margaret L. McNairy
- , Breige A. McNulty
- , Sounnia Mediene Benchekor
- , Jurate Medzioniene
- , Parinaz Mehdipour
- , Kirsten Mehlig
- , Amir Houshang Mehrparvar
- , Aline Meirhaeghe
- , Jørgen Meisfjord
- , Christa Meisinger
- , Jesus D. Melgarejo
- , Marina Melkumova

- , João Mello
- , Fabián Méndez
- , Carlos O. Mendivil
- , Ana Maria B. Menezes
- , Geetha R. Menon
- , Gert B. M. Mensink
- , Maria Teresa Menzano
- , Indrapal I. Meshram
- , Diane T. Meto
- , Jie Mi
- , Kim F. Michaelsen
- , Nathalie Michels
- , Kairit Mikkel
- , Karolina Miłkowska
- , Jody C. Miller
- , Olga Milushkina
- , Cláudia S. Minderico
- , G. K. Mini
- , Juan Francisco Miquel
- , J. Jaime Miranda
- , Mohammad Reza Mirjalili
- , Daphne Mirkopoulou
- , Erkin Mirrakhimov
- , Marjeta Mišigoj-Duraković
- , Antonio Mistretta
- , Veronica Mocanu
- , Pietro A. Modesti
- , Sahar Saeedi Moghaddam
- , Bahram Mohajer
- , Mostafa K. Mohamed
- , Shukri F. Mohamed
- , Kazem Mohammad
- , Mohammad Reza Mohammadi
- , Zahra Mohammadi
- , Noushin Mohammadifard
- , Reza Mohammadpourhodki
- , Viswanathan Mohan

- , Salim Mohanna
- , Muhammad Fadhli Mohd Yusoff
- , Iraj Mohebbi
- , Farnam Mohebi
- , Marie Moitry
- , Line T. Møllehave
- , Niels C. Møller
- , Dénes Molnár
- , Amirabbas Momenan
- , Charles K. Mondo
- , Roger A. Montenegro Mendoza
- , Eric Monterrubio-Flores
- , Kotsedi Daniel K. Monyeki
- , Jin Soo Moon
- , Mahmood Moosazadeh
- , Hermine T. Mopa
- , Farhad Moradpour
- , Leila B. Moreira
- , Alain Morejon
- , Luis A. Moreno
- , Francis Morey
- , Karen Morgan
- , Suzanne N. Morin
- , Erik Lykke Mortensen
- , George Moschonis
- , Alireza Moslem
- , Malgorzata Mossakowska
- , Aya Mostafa
- , Seyed-Ali Mostafavi
- , Anabela Mota-Pinto
- , Jorge Mota
- , Mohammad Esmaeel Motlagh
- , Jorge Motta
- , Marcos André Moura-dos-Santos
- , Yeva Movsesyan
- , Kelias P. Msyamboza
- , Thet Thet Mu

- , Magdalena Muc
- , Florian Muca
- , Boban Mugoša
- , Maria L. Muiesan
- , Martina Müller-Nurasyid
- , Thomas Münzel
- , Jaakko Mursu
- , Elaine M. Murtagh
- , Kamarul Imran Musa
- , Sanja Musić Milanović
- , Vera Musil
- , Geofrey Musinguzi
- , Muel Telo M. C. Muyer
- , Iraj Nabipour
- , Shohreh Naderimagham
- , Gabriele Nagel
- , Farid Najafi
- , Harunobu Nakamura
- , Hanna Nalecz
- , Jana Námešná
- , Ei Ei K. Nang
- , Vinay B. Nangia
- , Martin Nankap
- , Sameer Narake
- , Paola Nardone
- , Take Naseri
- , Matthias Nauck
- , William A. Neal
- , Azim Nejatizadeh
- , Chandini Nekkantti
- , Keiu Nelis
- , Ilona Nenko
- , Martin Neovius
- , Flavio Nervi
- , Tze Pin Ng
- , Chung T. Nguyen
- , Nguyen D. Nguyen

- , Quang Ngoc Nguyen
- , Michael Y. Ni
- , Rodica Nicolescu
- , Peng Nie
- , Ramfis E. Nieto-Martínez
- , Yury P. Nikitin
- , Guang Ning
- , Toshiharu Ninomiya
- , Nobuo Nishi
- , Sania Nishtar
- , Marianna Noale
- , Oscar A. Noboa
- , Helena Nogueira
- , Maria Nordendahl
- , Børge G. Nordestgaard
- , Davide Noto
- , Natalia Nowak-Szczepanska
- , Mohannad Al Nsour
- , Irfan Nuhoğlu
- , Baltazar Nunes
- , Eha Nurk
- , Fred Nuwaha
- , Moffat Nyirenda
- , Terence W. O'Neill
- , Dermot O'Reilly
- , Galina Obreja
- , Caleb Ochimana
- , Angélica M. Ochoa-Avilés
- , Eiji Oda
- , Augustine N. Odili
- , Kyungwon Oh
- , Kumiko Ohara
- , Claes Ohlsson
- , Ryutaro Ohtsuka
- , Örn Olafsson
- , Maria Teresa A. Olinto
- , Isabel O. Oliveira

- , Mohd Azahadi Omar
- , Saeed M. Omar
- , Altan Onat
- , Sok King Ong
- , N. Charlotte Onland-Moret
- , Lariane M. Ono
- , Pedro Ordunez
- , Rui Ornelas
- , Ana P. Ortiz
- , Pedro J. Ortiz
- , Merete Osler
- , Clive Osmond
- , Sergej M. Ostojic
- , Afshin Ostovar
- , Johanna A. Otero
- , Kim Overvad
- , Ellis Owusu-Dabo
- , Fred Michel Paccaud
- , Ioannis Pagkalos
- , Elena Pahomova
- , Karina Mary de Paiva
- , Andrzej Pająk
- , Alberto Palloni
- , Luigi Palmieri
- , Wen-Harn Pan
- , Songhomitra Panda-Jonas
- , Arvind Pandey
- , Francesco Panza
- , Mariela Paoli
- , Sousana K. Papadopoulou
- , Dimitrios Papandreou
- , Rossina G. Pareja
- , Soon-Woo Park
- , Suyeon Park
- , Winsome R. Parnell
- , Mahboubeh Parsaeian
- , Ionela M. Pascanu

- , Patrick Pasquet
- , Nikhil D. Patel
- , Marcos Pattussi
- , Halyna Pavlyshyn
- , Raimund Pechlaner
- , Ivan Pećin
- , Mangesh S. Pednekar
- , João M. Pedro
- , Nasheeta Peer
- , Sergio Viana Peixoto
- , Markku Peltonen
- , Alexandre C. Pereira
- , Marco A. Peres
- , Cynthia M. Pérez
- , Valentina Peterkova
- , Annette Peters
- , Astrid Petersmann
- , Janina Petkeviciene
- , Ausra Petrauskiene
- , Olga Petrovna Kovtun
- , Emanuela Pettenuzzo
- , Niloofar Peykari
- , Norbert Pfeiffer
- , Modou Cheyassin Phall
- , Son Thai Pham
- , Rafael N. Pichardo
- , Daniela Pierannunzio
- , Iris Pigeot
- , Hynek Pikhart
- , Aida Pilav
- , Lorenza Pilotto
- , Francesco Pistelli
- , Freda Pitakaka
- , Aleksandra Piwonska
- , Andreia N. Pizarro
- , Pedro Plans-Rubió
- , Alina G. Platonova

- , Bee Koon Poh
- , Hermann Pohlabeln
- , Nadija S. Polka
- , Raluca M. Pop
- , Stevo R. Popovic
- , Miquel Porta
- , Georg Posch
- , Anil Poudyal
- , Dimitrios Poulimeneas
- , Hamed Pouraram
- , Farhad Pourfarzi
- , Akram Pourshams
- , Hossein Poustchi
- , Rajendra Pradeepa
- , Alison J. Price
- , Jacqueline F. Price
- , Antonio Prista
- , Rui Providencia
- , Jardena J. Puder
- , Iveta Pudule
- , Maria Puiu
- , Margus Punab
- , Muhammed S. Qadir
- , Radwan F. Qasrawi
- , Mostafa Qorbani
- , Hedley K. Quintana
- , Pedro J. Quiroga-Padilla
- , Tran Quoc Bao
- , Stefan Rach
- , Ivana Radic
- , Ricardas Radisauskas
- , Salar Rahimikazerooni
- , Mahfuzar Rahman
- , Mahmudur Rahman
- , Olli Raitakari
- , Manu Raj
- , Tamerlan Rajabov

- , Sherali Rakhmatulloev
- , Ivo Rakovac
- , Sudha Ramachandra Rao
- , Ambady Ramachandran
- , Otim P. C. Ramadan
- , Virgílio V. Ramires
- , Jacqueline Ramke
- , Elisabete Ramos
- , Rafel Ramos
- , Lekhraj Rampal
- , Sanjay Rampal
- , Lalka S. Rangelova
- , Vayia Rarra
- , Ramon A. Rascon-Pacheco
- , Cassiano Ricardo Rech
- , Josep Redon
- , Paul Ferdinand M. Reganit
- , Valéria Regecová
- , Jane D. P. Renner
- , Judit A. Repasy
- , Cézane P. Reuter
- , Luis Revilla
- , Abbas Rezaianzadeh
- , Yeunsook Rho
- , Lourdes Ribas-Barba
- , Robespierre Ribeiro
- , Elio Riboli
- , Adrian Richter
- , Fernando Rigo
- , Attilio Rigotti
- , Natascia Rinaldo
- , Tobias F. Rinke de Wit
- , Ana I. Rito
- , Raphael M. Ritti-Dias
- , Juan A. Rivera
- , Reina G. Roa
- , Louise Robinson

- , Cynthia Robitaille
- , Romana Roccaldo
- , Daniela Rodrigues
- , Fernando Rodríguez-Artalejo
- , María del Cristo Rodriguez-Perez
- , Laura A. Rodríguez-Villamizar
- , Andrea Y. Rodríguez
- , Ulla Roggenbuck
- , Peter Rohloff
- , Fabian Rohner
- , Rosalba Rojas-Martinez
- , Nipa Rojroongwasinkul
- , Dora Romaguera
- , Elisabetta L. Romeo
- , Rafaela V. Rosario
- , Annika Rosengren
- , Ian Rouse
- , Vanessa Rouzier
- , Joel G. R. Roy
- , Maira H. Ruano
- , Adolfo Rubinstein
- , Frank J. Rühli
- , Jean-Bernard Ruidavets
- , Blanca Sandra Ruiz-Betancourt
- , Maria Ruiz-Castell
- , Emma Ruiz Moreno
- , Iuliia A. Rusakova
- , Kenisha Russell Jonsson
- , Paola Russo
- , Petra Rust
- , Marcin Rutkowski
- , Marge Saamel
- , Charumathi Sabanayagam
- , Hamideh Sabbaghi
- , Elena Sacchini
- , Harshpal S. Sachdev
- , Alireza Sadjadi

- , Ali Reza Safarpour
- , Sare Safi
- , Saeid Safiri
- , Mohammad Hossien Saghi
- , Olfa Saidi
- , Nader Saki
- , Sanja Šalaj
- , Benoit Salanave
- , Eduardo Salazar Martinez
- , Calogero Saleva
- , Diego Salmerón
- , Veikko Salomaa
- , Jukka T. Salonen
- , Massimo Salvetti
- , Margarita Samoutian
- , Jose Sánchez-Abanto
- , Inés Sánchez Rodríguez
- , Sandjaja
- , Susana Sans
- , Loreto Santa Marina
- , Ethel Santacruz
- , Diana A. Santos
- , Ina S. Santos
- , Lèlita C. Santos
- , Maria Paula Santos
- , Osvaldo Santos
- , Rute Santos
- , Tamara R. Santos
- , Jouko L. Saramies
- , Luis B. Sardinha
- , Nizal Sarrafzadegan
- , Thirunavukkarasu Sathish
- , Kai-Uwe Saum
- , Savvas Savva
- , Mathilde Savy
- , Norie Sawada
- , Mariana Sbaraini

- , Marcia Scazufca
- , Beatriz D. Schaan
- , Angelika Schaffrath Rosario
- , Herman Schargrodsky
- , Anja Schienkiewitz
- , Karin Schindler
- , Sabine Schipf
- , Carsten O. Schmidt
- , Ida Maria Schmidt
- , Andrea Schneider
- , Peter Schnohr
- , Ben Schöttker
- , Sara Schramm
- , Stine Schramm
- , Helmut Schröder
- , Constance Schultsz
- , Matthias B. Schulze
- , Aletta E. Schutte
- , Sylvain Sebert
- , Moslem Sedaghattalab
- , Rusidah Selamat
- , Vedrana Sember
- , Abhijit Sen
- , Idowu O. Senbanjo
- , Sadaf G. Sepanlou
- , Guillermo Sequera
- , Luis Serra-Majem
- , Jennifer Servais
- , Ľudmila Ševčíková
- , Svetlana Shalnova
- , Teresa Shamah-Levy
- , Seyed Morteza Shamshirgaran
- , Coimbatore Subramaniam Shanthirani
- , Maryam Sharafkhah
- , Sanjib K. Sharma
- , Jonathan E. Shaw
- , Amaneh Shayanrad

- , Ali Akbar Shayesteh
- , Lela Shengelia
- , Zumin Shi
- , Kenji Shibuya
- , Hana Shimizu-Furusawa
- , Tal Shimony
- , Rahman Shiri
- , Namuna Shrestha
- , Khairil Si-Ramlee
- , Alfonso Siani
- , Rosalynn Siantar
- , Abla M. Sibai
- , Labros S. Sidossis
- , Natalia Silitrari
- , Antonio M. Silva
- , Caroline Ramos de Moura Silva
- , Diego Augusto Santos Silva
- , Kelly S. Silva
- , Xueling Sim
- , Mary Simon
- , Judith Simons
- , Leon A. Simons
- , Agneta Sjöberg
- , Michael Sjöström
- , Natalia A. Skoblina
- , Gry Skodje
- , Tatyana Slazhnyova
- , Jolanta Slowikowska-Hilczer
- , Przemysław Slusarczyk
- , Liam Smeeth
- , Hung-Kwan So
- , Fernanda Cunha Soares
- , Grzegorz Sobek
- , Eugène Sobngwi
- , Morten Sodemann
- , Stefan Söderberg
- , Moesijanti Y. E. Soekatri

- , Agustinus Soemantri
- , Reecha Sofat
- , Vincenzo Solfrizzi
- , Mohammad Hossein Somi
- , Emily Sonestedt
- , Yi Song
- , Sajid Soofi
- , Thorkild I. A. Sørensen
- , Elin P. Sørgjerd
- , Charles Sossa Jérome
- , Victoria E. Soto-Rojas
- , Aïcha Soumaré
- , Alfonso Sousa-Poza
- , Slavica Sovic
- , Bente Sparboe-Nilsen
- , Karen Sparrenberger
- , Phoebe R. Spencer
- , Angela Spinelli
- , Igor Spiroski
- , Jan A. Staessen
- , Hanspeter Stamm
- , Kaspar Staub
- , Bill Stavreski
- , Jostein Steene-Johannessen
- , Peter Stehle
- , Aryeh D. Stein
- , George S. Stergiou
- , Jochanan Stessman
- , Ranko Stevanović
- , Jutta Stieber
- , Doris Stöckl
- , Jakub Stokwiszewski
- , Ekaterina Stoyanova
- , Gareth Stratton
- , Karien Stronks
- , Maria Wany Strufaldi
- , Lela Sturua

- , Ramón Suárez-Medina
- , Machi Suka
- , Chien-An Sun
- , Liang Sun
- , Johan Sundström
- , Yn-Tz Sung
- , Jordi Sunyer
- , Paibul Suriyawongpaisal
- , Nabil William G. Sweis
- , Boyd A. Swinburn
- , Rody G. Sy
- , René Charles Sylva
- , Moyses Szklo
- , Lucjan Szponar
- , Lorraine Tabone
- , E. Shyong Tai
- , Konstantinos D. Tambalis
- , Mari-Liis Tammesoo
- , Abdonas Tamosiunas
- , Eng Joo Tan
- , Xun Tang
- , Maya Tanrygulyyeva
- , Frank Tanser
- , Yong Tao
- , Mohammed Rasoul Tarawneh
- , Jakob Tarp
- , Carolina B. Tarqui-Mamani
- , Radka Taxová Braunerová
- , Anne Taylor
- , Julie Taylor
- , Félicité Tchibindat
- , Saskia Te Velde
- , William R. Tebar
- , Grethe S. Tell
- , Tania Tello
- , Yih Chung Tham
- , K. R. Thankappan

- , Holger Theobald
- , Xenophon Theodoridis
- , Nihal Thomas
- , Barbara Thorand
- , Betina H. Thuesen
- , Ľubica Tichá
- , Erik J. Timmermans
- , Dwi H. Tjandrarini
- , Anne Tjonneland
- , Hanna K. Tolonen
- , Janne S. Tolstrup
- , Murat Topbas
- , Roman Topór-Mądry
- , Liv Elin Torheim
- , María José Tormo
- , Michael J. Tornaritis
- , Maties Torrent
- , Laura Torres-Collado
- , Stefania Toselli
- , Giota Touloumi
- , Pierre Traissac
- , Thi Tuyet-Hanh Tran
- , Mark S. Tremblay
- , Areti Triantafyllou
- , Dimitrios Trichopoulos
- , Antonia Trichopoulou
- , Oanh T. H. Trinh
- , Atul Trivedi
- , Yu-Hsiang Tsao
- , Lechaba Tshepo
- , Maria Tsigga
- , Panagiotis Tsintavis
- , Shoichiro Tsugane
- , John Tuitele
- , Azaliia M. Tuliakova
- , Marshall K. Tulloch-Reid
- , Fikru Tullu

- , Tomi-Pekka Tuomainen
- , Jaakko Tuomilehto
- , Maria L. Turley
- , Gilad Twig
- , Per Tynelius
- , Evangelia Tzala
- , Themistoklis Tzotzas
- , Christophe Tzourio
- , Peter Ueda
- , Eunice Ugel
- , Flora A. M. Ukoli
- , Hanno Ulmer
- , Belgin Unal
- , Zhamyila Usupova
- , Hannu M. T. Uusitalo
- , Nalan Uysal
- , Justina Vaitkeviciute
- , Gonzalo Valdivia
- , Susana Vale
- , Damaskini Valvi
- , Rob M. van Dam
- , Bert-Jan van den Born
- , Johan Van der Heyden
- , Yvonne T. van der Schouw
- , Koen Van Herck
- , Wendy Van Lippevelde
- , Hoang Van Minh
- , Natasja M. Van Schoor
- , Irene G. M. van Valkengoed
- , Dirk Vanderschueren
- , Diego Vanuzzo
- , Anette Varbo
- , Gregorio Varela-Moreiras
- , Luz Nayibe Vargas
- , Patricia Varona-Pérez
- , Senthil K. Vasan
- , Daniel G. Vasques

- , Tomas Vega
- , Toomas Veidebaum
- , Gustavo Velasquez-Melendez
- , Biruta Velika
- , Maïté Verloigne
- , Giovanni Veronesi
- , W. M. Monique Verschuren
- , Cesar G. Victora
- , Giovanni Viegi
- , Lucie Viet
- , Frøydis N. Vik
- , Monica Vilar
- , Salvador Villalpando
- , Jesus Vioque
- , Jyrki K. Virtanen
- , Sophie Visvikis-Siest
- , Bharathi Viswanathan
- , Mihaela Vladulescu
- , Tiina Vlasoff
- , Dorja Vocanec
- , Peter Vollenweider
- , Henry Völzke
- , Ari Voutilainen
- , Martine Vrijheid
- , Tanja G. M. Vrijkotte
- , Alisha N. Wade
- , Thomas Waldhör
- , Janette Walton
- , Elvis O. A. Wambiya
- , Wan Mohamad Wan Bebakar
- , Wan Nazaimoon Wan Mohamud
- , Rildo de Souza Wanderley Júnior
- , Ming-Dong Wang
- , Ningli Wang
- , Qian Wang
- , Xiangjun Wang
- , Ya Xing Wang

- , Ying-Wei Wang
- , S. Goya Wannamethee
- , Nicholas Wareham
- , Adelheid Weber
- , Karen Webster-Kerr
- , Niels Wedderkopp
- , Daniel Weghuber
- , Wenbin Wei
- , Aneta Weres
- , Bo Werner
- , Leo D. Westbury
- , Peter H. Whincup
- , Kremlin Wickramasinghe
- , Kurt Widhalm
- , Indah S. Widyahening
- , Andrzej Więcek
- , Philipp S. Wild
- , Rainford J. Wilks
- , Johann Willeit
- , Peter Willeit
- , Julianne Williams
- , Tom Wilsgaard
- , Rusek Wojciech
- , Bogdan Wojtyniak
- , Kathrin Wolf
- , Roy A. Wong-McClure
- , Andrew Wong
- , Emily B. Wong
- , Jyh Eiin Wong
- , Tien Yin Wong
- , Jean Woo
- , Mark Woodward
- , Frederick C. Wu
- , Hon-Yen Wu
- , Jianfeng Wu
- , Li Juan Wu
- , Shouling Wu

- , Justyna Wyszyńska
- , Haiquan Xu
- , Liang Xu
- , Nor Azwany Yaacob
- , Uruwan Yamborisut
- , Weili Yan
- , Ling Yang
- , Xiaoguang Yang
- , Yang Yang
- , Nazan Yardim
- , Tabara Yasuharu
- , Martha Yépez García
- , Panayiotis K. Yiallouros
- , Agneta Yngve
- , Moein Yoosefi
- , Akihiro Yoshihara
- , Qi Sheng You
- , San-Lin You
- , Novie O. Younger-Coleman
- , Yu-Ling Yu
- , Yunjiang Yu
- , Safiah Md Yusof
- , Ahmad Faudzi Yusoff
- , Luciana Zaccagni
- , Vassilis Zafiropulos
- , Ahmad A. Zainuddin
- , Seyed Rasoul Zakavi
- , Farhad Zamani
- , Sabina Zambon
- , Antonis Zampelas
- , Hana Zamrazilová
- , Maria Elisa Zapata
- , Abdul Hamid Zargar
- , Ko Ko Zaw
- , Ayman A. Zayed
- , Tomasz Zdrojewski
- , Magdalena Żegleń

- , Kristyna Zejglicova
- , Tajana Zeljkovic Vrkic
- , Yi Zeng
- , Luxia Zhang
- , Zhen-Yu Zhang
- , Dong Zhao
- , Ming-Hui Zhao
- , Wenhua Zhao
- , Yanitsa V. Zhecheva
- , Shiqi Zhen
- , Wei Zheng
- , Yingfeng Zheng
- , Bekbolat Zholdin
- , Maigeng Zhou
- , Dan Zhu
- , Marie Zins
- , Emanuel Zitt
- , Yanina Zocalo
- , Nada Zoghlami
- , Julio Zuñiga Cisneros
- , Monika Zuziak
- , Zulfiqar A. Bhutta
- , Robert E. Black
- & Majid Ezzati

Contributions

A.M., B.Z., A.R.M., H.B. and R.K.S. led the data collection and management. A.M., B.Z., A.R.M., H.B., C.J.P., J.E.B. and M.E. developed the statistical method. A.M., B.Z., A.R.M. and H.B. coded the statistical method. A.M. conducted analyses and prepared results. The other authors contributed to study design, and collected, reanalysed, checked and pooled data. M.E., A.M., B.Z., A.R.M. and H.B. wrote the first draft of the report. All other authors commented on the draft report.

Corresponding author

Correspondence to Majid Ezzati.

Ethics declarations

Competing interests

M.E. reports a charitable grant from the AstraZeneca Young Health Programme.

Peer review

Peer review information

Nature thanks Sam Harper, Jessica Jones-Smith and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Number of data sources used in the analysis, by country

.

Extended Data Fig. 2 Urban-rural height difference in 2020 and change from 1990 to 2020.

The top two maps show the urban-rural difference in age-standardised mean height in 2020 for girls and boys resepectively. A positive number shows

higher urban mean height and a negative number shows higher rural mean height. The bottom two maps show the change from 1990 to 2020. The density plot below each map shows the distribution of estimates across countries. The top two scatter plots show the urban-rural difference in age-standardised mean height in relation to the uncertainty of the difference measured by posterior s.d. The bottom two scatter plots show the change from 1990 to 2020 in urban-rural difference in mean height in relation to the uncertainty of the change measured by posterior s.d. Each point in the scatter plots shows one country. Shaded areas approximately show the posterior probability (PP) of a true difference (top two scatter plots) and of a true increase or decrease in difference (bottom two scatter plots). See Extended Data Fig. 8 for PPs of the urban-rural difference in age-standardised mean height and its change. See Supplementary Fig. 7 for results at ages 5, 10, 15 and 19 years. We did not estimate the difference between rural and urban height for countries classified as entirely urban (Bermuda, Kuwait, Nauru and Singapore) or entirely rural (Tokelau), as indicated in grey.

Extended Data Fig. 3 Urban-rural body-mass index (BMI) difference in 2020 and change from 1990 to 2020.

See Extended Data Fig. 2 caption for descriptions of the contents of the figure and for definitions. See Extended Data Fig. 9 for PP of the urban-rural difference in age-standardised mean BMI and its change. See Supplementary Fig. 8 for results at ages 5, 10, 15 and 19 years. We did not estimate the difference between rural and urban BMI for countries classified as entirely urban (Bermuda, Kuwait, Nauru and Singapore) or entirely rural (Tokelau), as indicated in grey.

Extended Data Fig. 4 Posterior probability of increase in mean height in urban and rural areas from 1990 to 2020.

The maps show the PP that the age-standardised mean height increased from 1990 to 2020. The PP of a decrease is one minus that of an increase. If an increase in mean height is statistically indistinguishable from a decrease, the PP is 0.50. PPs closer to 0.50 indicate more uncertainty, those towards 1

indicate more certainty of an increase, and those towards 0 indicate more certainty of a decrease. We did not estimate PP for change in mean rural height for countries classified as entirely urban (Bermuda, Kuwait, Nauru and Singapore) or change in mean urban height for countries classified as entirely rural (Tokelau), as indicated in grey.

Extended Data Fig. 5 Posterior probability of increase in mean body-mass index (BMI) in urban and rural areas from 1990 to 2020.

The maps show the posterior probability (PP) that the age-standardised mean BMI increased from 1990 to 2020. The PP of a decrease is one minus that of an increase. We did not estimate PP for change in mean rural BMI in countries classified as entirely urban (Bermuda, Kuwait, Nauru and Singapore) or change in mean urban BMI in countries classified as entirely rural (Tokelau), as indicated in grey.

Extended Data Fig. 6 Trends in body-mass index (BMI) by place of residence for children, adolescents and young adults for females.

The figure shows trends in mean BMI at ages five and 19 years, and in age-standardised mean BMI for young adults (20–29 years and 30–39 years) for females. Shaded areas show the 95% CrIs. Trend for young adults were estimated using a model similar to the one described in Methods, where BMI-age patterns were allowed to vary flexibly via a cubic spline function without knots.

Extended Data Fig. 7 Trends in body-mass index (BMI) by place of residence for children, adolescents and young adults for males.

The figure shows trends in mean BMI at ages five and 19 years, and in age-standardised mean BMI for young adults (20–29 years and 30–39 years) for males. See Extended Data Fig. 6 caption for description of figure contents.

Extended Data Fig. 8 Posterior probability of urban-rural height difference in 2020 and its increase from 1990 to 2020.

The maps show the posterior probability (PP) that age-standardised mean height in 2020 in urban areas was higher than in rural areas (left-hand panels), and the PP that the urban-rural difference in age-standardised mean height increased from 1990 to 2020 (right-hand panels). For 2020, if estimated age-standardised mean urban height is statistically indistinguishable from rural height, the PP is 0.50. PPs closer to 0.50 indicate more uncertainty, those towards 1 indicate more certainty of urban children being taller, and those towards 0 indicate more certainty of rural being taller. For change, if an increase in urban-rural difference in mean height is statistically indistinguishable from a decrease, the PP is 0.50. PPs closer to 0.50 indicate more uncertainty, those towards 1 indicate more certainty of an increase in the urban-rural height difference, and those towards 0 indicate more certainty of a decrease. We did not estimate the PP for differences between rural and urban height for countries classified as entirely urban (Bermuda, Kuwait, Nauru and Singapore) or entirely rural (Tokelau), as indicated in grey.

Extended Data Fig. 9 Posterior probability of urban-rural body-mass index (BMI) difference in 2020 and its increase from 1990 to 2020.

The maps show the posterior probability (PP) that age-standardised mean BMI in 2020 in urban areas was higher than in rural areas (left-hand panels), and the PP that the urban-rural difference in mean BMI increased from 1990 to 2020 (right-hand panels). We did not estimate the PP for differences between rural and urban BMI for countries classified as entirely urban (Bermuda, Kuwait, Nauru and Singapore) or entirely rural (Tokelau), as indicated in grey.

Supplementary information

Supplementary Information

Supplementary Tables 1–4, Supplementary Figs. 1–8 and Supplementary References; see contents page for details.

Reporting Summary

Peer Review File

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

NCD Risk Factor Collaboration (NCD-RisC). Diminishing benefits of urban living for children and adolescents' growth and development. *Nature* **615**, 874–883 (2023). https://doi.org/10.1038/s41586-023-05772-8

- Received: 27 May 2022
- Accepted: 30 January 2023

- Published: 29 March 2023
- Issue Date: 30 March 2023
- DOI: https://doi.org/10.1038/s41586-023-05772-8

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

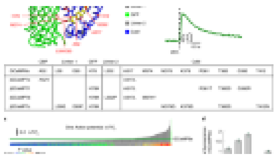
Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

Health benefits of growing up in cities have diminished globally

Nature Research Briefing 29 Mar 2023

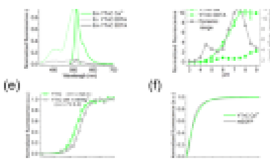
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05772-8



High-performance calcium sensors for imaging activity in neuronal populations and microcompartments

17 June 2019

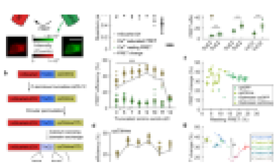
Hod Dana, Yi Sun, … Douglas S. Kim



NTnC-like genetically encoded calcium indicator with a positive and enhanced response and fast kinetics

15 October 2018

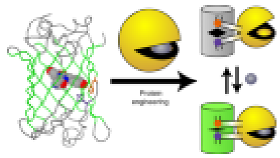
Natalia V. Barykina, Danila A. Doronin, … Fedor V. Subach



Rational engineering of ratiometric calcium sensors with bright green and red fluorescent proteins

29 July 2021

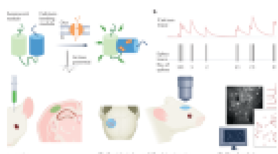
Diming Zhang, Emily Redington & Yiyang Gong



[Structure- and mechanism-guided design of single fluorescent protein-based biosensors]

08 February 2021

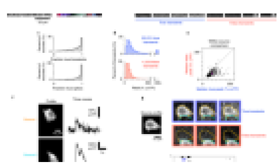
Yusuke Nasu, Yi Shen, … Robert E. Campbell



[Two-photon calcium imaging of neuronal activity]

01 September 2022

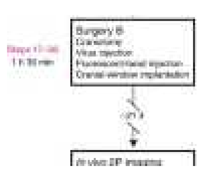
Christine Grienberger, Andrea Giovannucci, … Carlos Portera-Cailliau



[Detecting and correcting false transients in calcium imaging]

28 March 2022

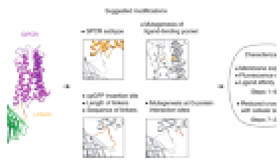
Jeffrey L. Gauthier, Sue Ann Koay, … Adam S. Charles



High-yield in vitro recordings from neurons functionally characterized in vivo

10 May 2018

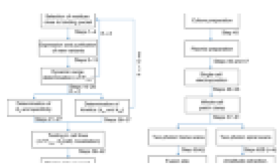
Simon Weiler, Joel Bauer, … Volker Scheuss



Imaging neuromodulators with high spatiotemporal resolution using genetically encoded indicators

15 November 2019

Tommaso Patriarchi, Jounhong Ryan Cho, … Lin Tian



High-speed imaging of glutamate release with genetically encoded sensors

15 April 2019

Céline D. Dürst, J. Simon Wiegert, … Thomas G. Oertner

- Article
- Open Access
- Published: 15 March 2023

Fast and sensitive GCaMP calcium indicators for imaging neural populations

- Yan Zhang,
- Márton Rózsa,
- Yajie Liang,
- Daniel Bushey,
- Ziqiang Wei,
- Jihong Zheng,
- Daniel Reep,
- Gerard Joey Broussard,
- Arthur Tsang,
- Getahun Tsegaye,
- Sujatha Narayan,
- Christopher J. Obara,
- Jing-Xuan Lim,
- Ronak Patel,
- Rongwei Zhang,
- Misha B. Ahrens,
- Glenn C. Turner ,
- Samuel S.-H. Wang ,
- Wyatt L. Korff,
- Eric R. Schreiter,
- Karel Svoboda ,
- Jeremy P. Hasseman ,
- Ilya Kolb &
- …
- Loren L. Looger

Show authors

Nature **volume 615**, pages 884–891 (2023)

- 15k Accesses
- 151 Altmetric
- Metrics details

Abstract

Calcium imaging with protein-based indicators[1,2] is widely used to follow neural activity in intact nervous systems, but current protein sensors report neural activity at timescales much slower than electrical signalling and are limited by trade-offs between sensitivity and kinetics. Here we used large-scale screening and structure-guided mutagenesis to develop and optimize several fast and sensitive GCaMP-type indicators[3,4,5,6,7,8]. The resulting 'jGCaMP8' sensors, based on the calcium-binding protein calmodulin and a fragment of endothelial nitric oxide synthase, have ultra-fast kinetics (half-rise times of 2 ms) and the highest sensitivity for neural activity reported for a protein-based calcium sensor. jGCaMP8 sensors will allow tracking of large populations of neurons on timescales relevant to neural computation.

Main

Measurement of Ca^{2+}-dependent fluorescence using genetically encoded calcium indicators (GECIs)[1,2] is a standard method for tracking neural activity in defined neurons and neural networks[9,10]. Recent advances have been driven by engineered GECIs with higher sensitivity[3,4,5,6,7,8], which in turn have stimulated the development of new methods for in vivo microscopy[11,12,13]. In particular, the green fluorescent protein (GFP)-based GCaMP sensors[2,3,5,6] have been iteratively engineered to enhance the signal-to-noise ratio (SNR) for detecting Ca^{2+} entering neurons during neural activity. The GCaMP6 (ref. [5]) and jGCaMP7 (ref. [6]) sensors enable detection of single action potentials (APs) under favourable conditions and are often used to monitor the activity of large groups of neurons using two-photon microscopy or wide-field fluorescence imaging[11,12]. They have also been used to measure activity-induced calcium changes in small synaptic compartments such as dendritic spines[5] and axons[14].

Electrical signals propagate through neural circuits over timescales of milliseconds. Determining how the activity of one set of neurons influences another and ultimately animal behaviour requires tracking activity on concomitant timescales. In many neuron types, APs produce large (more than 1 μM) and rapid (rise time of less than 1 ms) increases in cytoplasmic-free calcium[15]. Calcium ions can activate fluorescent calcium indicators very rapidly. For example, millisecond-timescale detection of APs has been demonstrated with synthetic fluorescent calcium indicators[16,17,18]. However, the kinetics of GECI fluorescence changes are relatively slow and limited by sensor biophysics downstream of calcium binding[19]. In response to single APs in pyramidal neurons, most widely used GCaMPs have fluorescent half-rise times on the order of 100 ms (refs. [3,5,6,19,20]). Consequently, GCaMPs are often used to map relatively static representations of neural information, rather than tracking the rich dynamics in neural circuits[21,22].

Previous attempts to improve GCaMP kinetics have been only partially successful. Among the GCaMP6 (ref. [5]) and jGCaMP7 (ref. [6]) indicators, the fast (f) variants were optimized for kinetics. They have faster rise and decay times, but with reduced sensitivity compared with their slower relatives (sensitive (s) variants). Generally, attempts to improve SNR are associated with a slowing of kinetics[5,6,20]. The mechanisms underlying this trade-off are not simply due to changes in affinity for Ca^{2+}. For example, the kinetics are sensitive to mutations at the interface between calmodulin (CaM) and the CaM-binding peptide RS20 (derived from myosin light chain kinase[23]), far from the Ca^{2+}-binding EF hands on CaM[5,20]. Recently, the RS20 peptide has been swapped for the peptide from CaM-dependent kinase kinase CaMKK-α/β (ckkap peptide) in some GECIs. The resulting XCaMP and R-CaMP2 sensors provide faster kinetics[7,8], confirming that calcium-dependent interactions between CaM and the CaM-binding peptide help to determine sensor kinetics[5].

Here we present GCaMP sensors with improved kinetics without compromising sensitivity or brightness. jGCaMP8 sensors include: jGCaMP8s (fast rise, slow decay and sensitive), jGCaMP8f (fast rise and fast decay) and jGCaMP8m (fast rise and medium decay). All jGCaMP8 sensors have nearly tenfold-faster fluorescence rise times than previous GCaMPs and can track individual spikes in neurons with spike rates up to

50 Hz. jGCaMP8 sensors are also more linear than previous GCaMPs, allowing robust deconvolution for spike extraction. The jGCaMP8 sensors were tested in vivo in mice and flies and provide better performance across multiple metrics relevant to imaging neural populations in vivo.

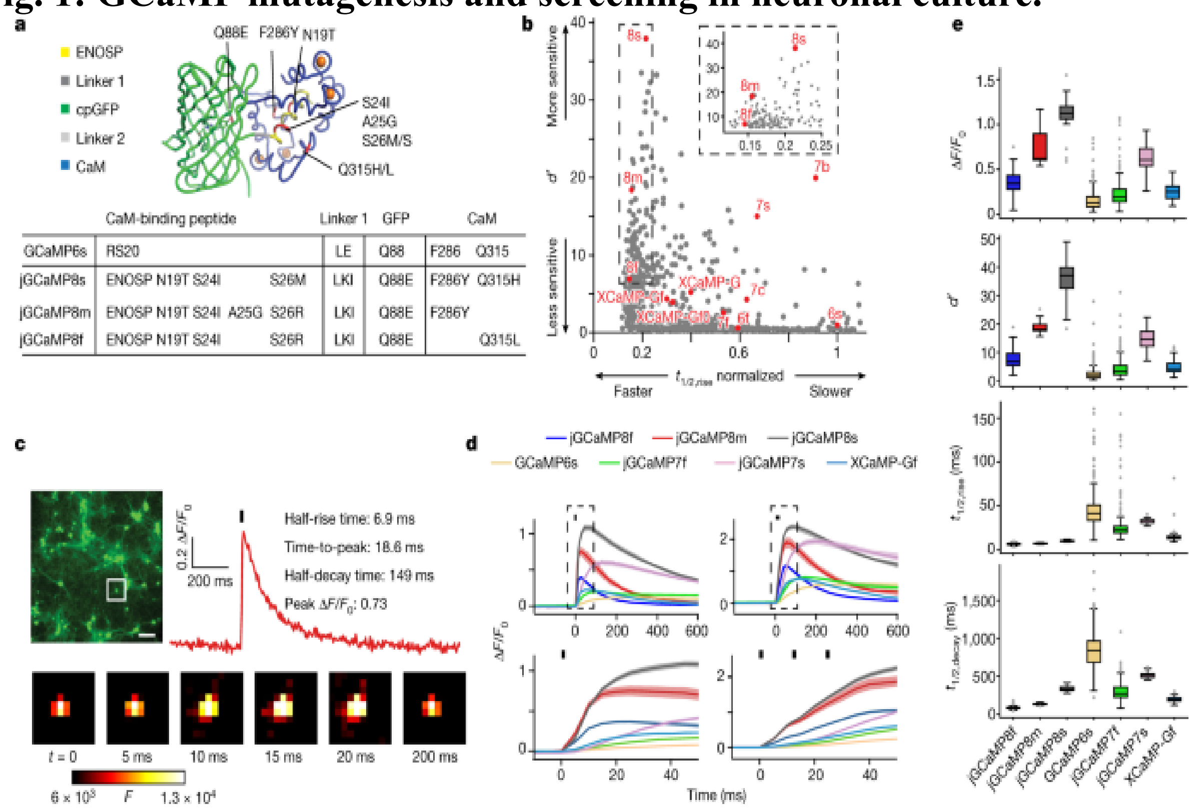
Sensor design and optimization

Various CaM-binding peptides (Extended Data Table 1) were cloned into GCaMP6s, replacing RS20. Basic sensor properties were measured in bacterially purified protein, including fluorescence increase ($(F_{sat} - F_{apo})/F_{apo}$) upon saturating calcium binding, half-decay time ($t_{1/2,decay}$) of fluorescence after calcium removal, apparent binding constant K_d, Hill coefficient (cooperativity) and apparent brightness. On the basis of these measurements (Extended Data Table 1), we prioritized variants based on peptides from endothelial nitric oxide synthase (ENOSP) and death-associated protein kinase 1 (DAPKP) for optimization (Methods). The two linkers[4] were systematically mutated, and sensors were screened for high calcium-dependent signal change, while retaining short $t_{1/2,decay}$, in purified protein. Thirty-five promising sensors were then tested in response to APs in cultured neurons in 96-well plates[24] (Methods). APs produce essentially instantaneous increases in calcium[15] and are therefore ideal to screen for GECIs with fast kinetics[25]. Fluorescence changes were extracted from multiple single neurons per well. Sensors were evaluated based on several properties (Extended Data Table 1 and Supplementary Table 1): baseline brightness (F_0); fluorescence change ($\Delta F/F_0 = (F - F_0)/F_0$) in response to a single AP (1AP $\Delta F/F_0$); fluorescence response to a saturating high-frequency train of 160 APs (reflective of total dynamic range); SNR quantified as the sensitivity index d' (ref. [26]) and kinetics (half-rise time ($t_{1/2,rise}$) and $t_{1/2,decay}$). Sensors based on DAPKP showed fast half-decay time and high sensitivity compared with jGCaMP7f, but with slow half-rise times. Sensors with ENOSP had similar sensitivity and substantially faster half-rise and half-decay times than jGCaMP7f.

We prioritized ENOSP-based sensors for further optimization, starting with variant jGCaMP8.410.80, which has a 1.8-fold faster 1AP half-rise time

(1AP $t_{1/2,\mathrm{rise}}$) and a 4.4-fold faster 1AP half-decay time (1AP $t_{1/2,\mathrm{decay}}$) than jGCaMP7f, with similar resting brightness, dynamic range and sensitivity. We solved the crystal structure of jGCaMP8.410.80 (Fig. 1a, Extended Data Table 2 and Extended Data Fig. 1). Guided by the structure, we targeted interface sites (Extended Data Fig. 1c) for site-saturation mutagenesis and tested variants in cultured neurons for sensitivity and fast kinetics in response to APs. Multiple single mutations, particularly residues near the ENOSP C terminus and the interface, improved properties (Supplementary Table 1). Beneficial point mutations were combined in subsequent rounds of screening[5].

Fig. 1: GCaMP mutagenesis and screening in neuronal culture.



a, jGCaMP8 (variant 8.410.80) structure and mutations in different jGCaMP8 variants relative to GCaMP6s (top). ENOSP, linker 1 (ENOSP–cpGFP), linker 2 (cpGFP–CaM), cpGFP, CaM, mutated sites (red) and Ca^{2+} ions (orange) are shown. Mutations for each jGCaMP8 variant (bottom table) are also displayed. **b**, Sensitivity (d') and rise kinetics ($t_{1/2,\mathrm{rise}}$) for jGCaMP8 variants. The *x* axis is normalized to GCaMP6s. GCaMP6, jGCaMP7, jGCaMP8 and XCaMP are highlighted in red. Mutants with

normalized $t_{1/2,rise}$ > 1.1 are not shown. The inset shows a zoomed in view on the jGCaMP8 series. Complete multi-parameter scatterplots are available as an interactive Binder notebook (Methods). **c**, Screening in neurons. Field stimulation of jGCaMP8m-expressing cultured neurons (top left), a fluorescence trace (1AP) (top right) and single frames of *F* corresponding to the box in the image (bottom) are shown. Scale bar, 100 µm. **d**, Responses to 1AP (black bar; top left) and 3AP (black bars; top right). Zoomed-in insets from the top panel (dashed boxes) to highlight rise kinetics are also shown (bottom). Solid lines indicate the mean and the shaded area denotes s.e.m. (n = 48 wells and 1,696 neurons (jGCaMP8f), n = 11 wells and 496 neurons (jGCaMP8m), n = 24 wells and 1,183 neurons (jGCaMP8s), n = 283 wells and 8,700 neurons (GCaMP6s), n = 294 wells and 7,372 neurons (jGCaMP7f), n = 22 wells and 514 neurons (jGCaMP7s), and n = 69 wells and 1,305 neurons (XCaMP-Gf); overall statistics, n = 7 independent transfections, 38 96-well plates). Data shown represent a portion of the overall screened constructs in Supplementary Table 1. **e**, Responses to 1AP for jGCaMP8 indicators and comparison with GCaMP6s, jGCaMP7f, jGCaMP7s and XCaMP-Gf. Data and *n* values are the same as in **d**. For the box-and-whisker plots, the box indicates the median and 25–75th percentile range, and the whiskers indicate the shorter of 1.5 times the interquartile range or the extreme data point.

Screening in neurons covered 813 jGCaMP8 sensor variants (Supplementary Table 1), of which 647 (80%) produced detectable responses to 1AP (Extended Data Fig. 2). In addition, nine previously developed GECIs were included in the screen for comparison. Compared with jGCaMP7f, 1AP $t_{1/2,rise}$ and 1AP $t_{1/2,decay}$ were significantly shorter in 47% and 48% of variants, respectively. The 1AP $\Delta F/F_0$ was higher than jGCaMP7f in 19% of variants. Together, mutagenesis produced a large set of variants with significant improvement in kinetics and sensitivity to neural activity (Supplementary Table 1).

jGCaMP8 characterization

Three high-performing ‘jGCaMP8’ variants were selected for additional characterization (Fig. 1b–e, Extended Data Table 3 and Extended Data Fig.

3). jGCaMP8f (fast) exhibited 1AP $t_{1/2,\mathrm{rise}}$ of 6.6 ± 1.0 ms, more than threefold shorter than jGCaMP7f. jGCaMP8s (sensitive) exhibited the highest 1AP $\Delta F/F_0$ and 1AP d' of any construct measured. For jGCaMP8s, 1AP d' was approximately twice that of the most sensitive GECI to date, jGCaMP7s. jGCaMP8m (medium) is a compromise between sensitivity and kinetics: it exhibits 1AP d' comparable with jGCaMP7s, and kinetics comparable with jGCaMP8f, apart from a slower half-decay time (Fig. 1d,e and Extended Data Table 3). Overall, the jGCaMP8 series exhibited significant, multifold improvements across several parameters over previous GECIs (Fig. 1b and Extended Data Table 4).

We then compared the new jGCaMP8 sensors to the recent XCaMP series (the green XCaMP variants XCaMP-G, XCaMP-Gf and XCaMP-Gf0)[8]. The 1AP $\Delta F/F_0$ was significantly higher for all jGCaMP8 sensors; 1AP d' was significantly higher for jGCaMP8m and jGCaMP8s; and kinetics were significantly faster for jGCaMP8f than the XCaMP sensors (Extended Data Table 4). Baseline fluorescence of the jGCaMP8 series was similar to jGCaMP7f, and significantly higher than the XCaMP sensors (Extended Data Fig. 4a). Photobleaching was also similar between jGCaMP7f and the jGCaMP8 sensors (Extended Data Table 3 and Extended Data Fig. 4b). In equimolar purified protein, the 488-nm absorbance of XCaMP-Gf was approximately eight times lower than jGCaMP7f and the jGCaMP8 sensors in the Ca^{2+}-bound bright state, and the two-photon cross-section was also approximately eight times weaker (Extended Data Fig. 5). XCaMP has a much higher Ca^{2+}-bound pK_a (Extended Data Table 3), meaning a lower proportion of deprotonated bright fluorophore at physiological pH. This is consistent with its low extinction coefficient (Extended Data Table 3).

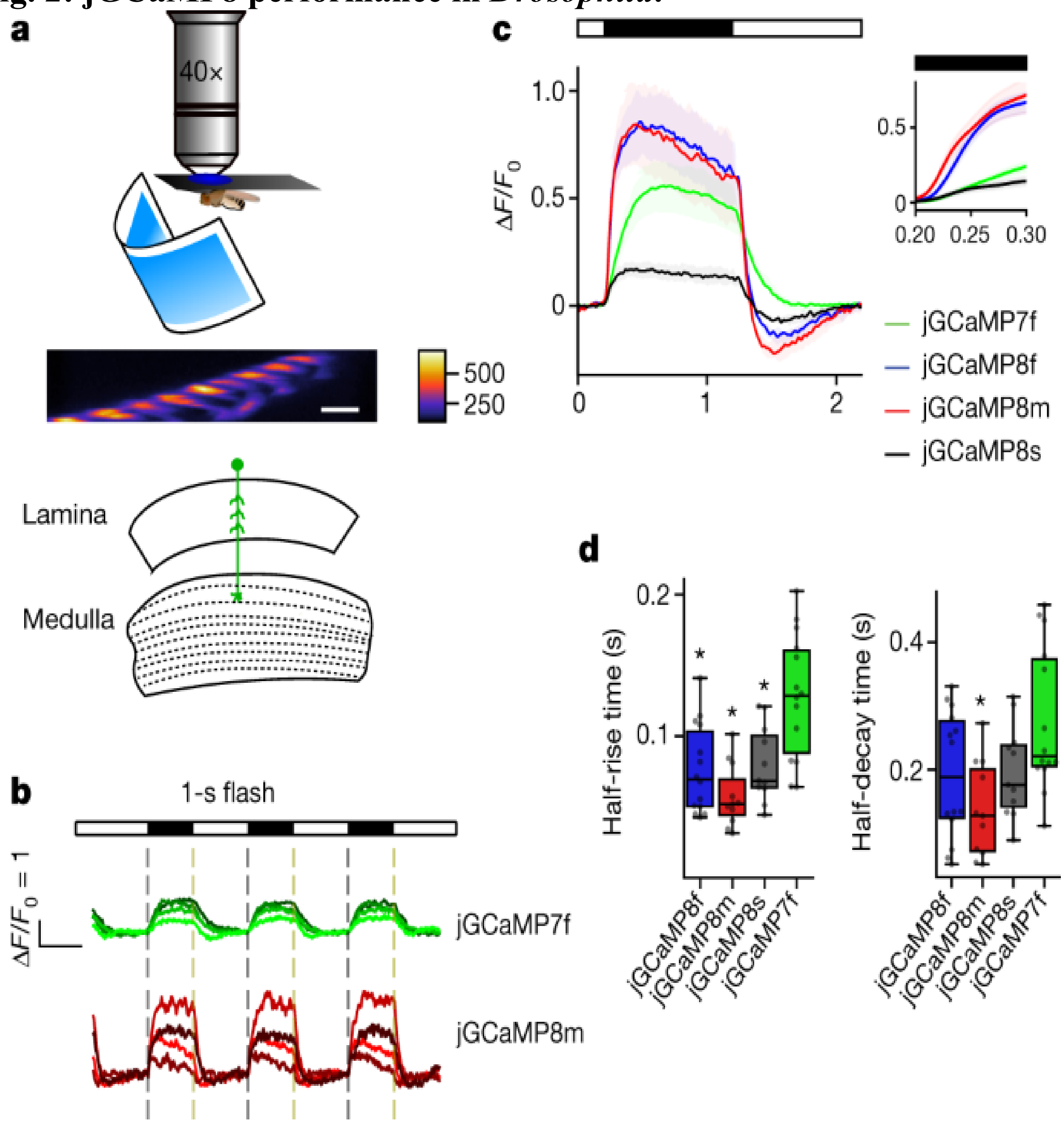
GECIs with linear fluorescence responses to AP trains provide a large effective dynamic range for quantifying spike rates and facilitate counting spikes within trains. In purified protein, Hill coefficients were lower for the jGCaMP8 variants (1.9–2.2) than jGCaMP7f (3.1) (Extended Data Table 3). We then tested GCaMP sensors with bursts (83 Hz) containing different numbers (1–40) of APs. Given their higher sensitivity to neural activity, jGCaMP8m and jGCaMP8s saturated at smaller numbers of spikes than the jGCaMP7 sensors. However, they behaved nearly linearly up to ten spikes (Extended Data Fig. 6). Finally, fluorescence recovery after photobleaching

revealed that the jGCaMP8 variants showed similar diffusion in neurons compared with previous GECIs[25] (Extended Data Fig. 7a–c) and independent of calcium (Extended Data Fig. 7d), suggesting that they do not have altered cellular interactions.

Imaging in larval and adult flies

jGCaMP8 responses to visual stimulation were measured in *Drosophila* laminar monopolar L2 neurons (Fig. 2a), which are part of the OFF-motion visual system[27]. These non-spiking neurons depolarize during light decrease and hyperpolarize during increase. Imaging was performed where L2 dendrites connect to columns in medulla layer 2. Fluorescence responses were first measured in multiple single neurons in response to 0.5-Hz light–dark flashes (Fig. 2b and Extended Data Fig. 8a). XCaMP-Gf, introduced using identical genetic strategies as the jGCaMP7 and jGCaMP8 sensors, was too dim to image (Extended Data Fig. 8b,c) and poorly expressed (Extended Data Fig. 9). At light–dark and dark–light transitions, all jGCaMP8 variants showed significantly faster rise, and jGCaMP8m showed faster decay, than jGCaMP7f (Fig. 2c,d). jGCaMP8m and jGCaMP8f also showed markedly larger fluorescence changes ($\Delta F/F_0$) than jGCaMP7f following light-on (Fig. 2b,c). All three jGCaMP8 indicators exhibited a negative off-response (Fig. 2c) after light-off (that is, hyperpolarization below baseline), consistent with previous electrophysiological[28] and voltage imaging experiments[29]. Flies were next subjected to light on–off stimulation at frequencies from 0.5 to 30 Hz. In power spectra of the fluorescence signal, jGCaMP8m and jGCaMP8f showed higher spectral density than jGCaMP8s across all frequencies, and higher than jGCaMP7f above 2 Hz (Extended Data Fig. 8d). Next, short dark flashes (duration of 4–25 ms) were shown to evaluate the impulse response of the sensors. jGCaMP8m and jGCaMP8f showed higher $\Delta F/F_0$ at all stimulus durations (Extended Data Fig. 8e, top). jGCaMP8m and jGCaMP8f provided markedly superior stimulus detection than jGCaMP7f and jGCaMP8s across all dark flash durations (Extended Data Fig. 8e, bottom). The jGCaMP8 variants were somewhat dimmer than jGCaMP7f because of lower expression (Extended Data Figs. 8b,c and 9) but were sufficiently bright to provide high SNR imaging.

Fig. 2: jGCaMP8 performance in *Drosophila*.



a, Schematic of the experiment. Fly with visual stimulus (top), fluorescence micrograph of L2 dendrites in medullar layer 2 (scale bar, 5 μm) (middle), and a schematic of the *Drosophila* visual system (bottom) are shown. **b**, $\Delta F/F_0$ response to a 0.5-Hz visual stimulation frequency from variants jGCaMP7f and jGCaMP8m. Individual traces show four representative individual animals per GECI (shading arbitrary). Light and dark periods are indicated by white and black bars above the graph. The error bands represent s.e.m. **c**, Mean $\Delta F/F_0$ response to 0.5-Hz stimulation. The solid line

indicates the mean and the shaded area denotes s.e.m. The dark period is represented by a black bar above the graph. The mean was calculated from eight trials per animal and then between animals. The inset compares the response from each variant at the onset of the dark period. **d**, Half-rise and half-decay times for responses in **c**. Half-rise: 128 ± 11 ms (jGCaMP7f), 76 ± 8 ms (jGCaMP8f), 58 ± 6 ms (jGCaMP8m) and 80 ± 8 ms (jGCaMP8s) (Kruskal–Wallis multiple-comparison test, $P = 2.9 \times 10^{-4}$; pairwise Dunn's comparison test with jGCaMP7f: $P = 3.1 \times 10^{-3}$ (jGCaMP8f), $P = 2.9 \times 10^{-5}$ (jGCaMP8m) and $P = 1.3 \times 10^{-2}$ (jGCaMP8s)). Half-decay times: 277 ± 29 ms (jGCaMP7f), 192 ± 26 ms (jGCaMP8f), 137 ± 21 ms (jGCaMP8m) and 198 ± 21 ms (jGCaMP8s) (Kruskal–Wallis multiple-comparison test, $P = 2.4 \times 10^{-2}$; pairwise Dunn's comparison test: $P = 1.1 \times 10^{-1}$ (jGCaMP8f), $P = 2.2 \times 10^{-3}$ (jGCaMP8m) and $P = 1.8 \times 10^{-1}$ (jGCaMP8s)). $^{*}P < 0.05$. Total n of flies tested for each variant in **c** and **d**: 14 (jGCaMP7f), 11 (jGCaMP8s), 11 (jGCaMP8m) and 14 (jGCaMP8f). For the box-and-whisker plots, the box indicates the median and 25–75th percentile range, and the whiskers indicate the shorter of 1.5 times the interquartile range or the extreme data point.

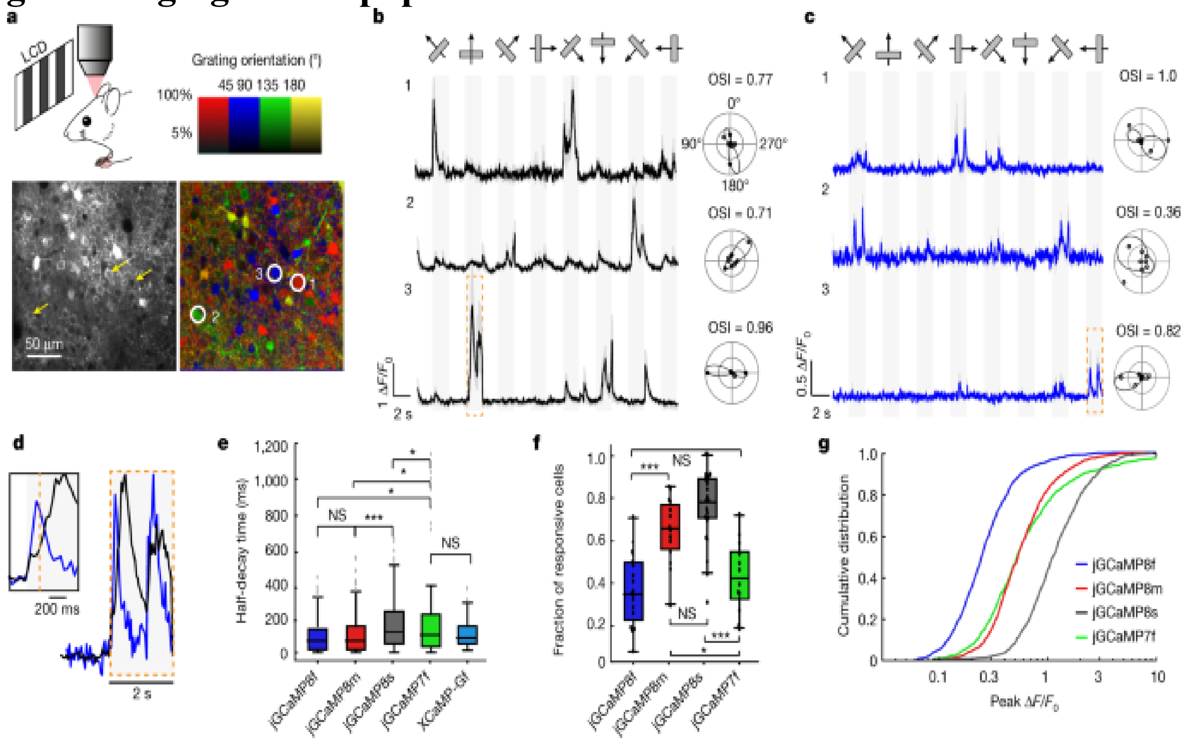
Next, we imaged jGCaMP8 responses at presynaptic boutons of the larval neuromuscular junction in response to electrical stimulation of motor axons[5] (Extended Data Fig. 10). jGCaMP8 variants showed large responses, with faster rise and decay times than jGCaMP7f (Extended Data Fig. 10b,d,e). The jGCaMP8 series detected individual stimuli better than jGCaMP7f at low frequencies and easily resolved spikes in 20-Hz stimulation trains (Extended Data Fig. 10h).

Imaging in the mouse visual cortex

We next tested the jGCaMP8 sensors in L2/3 pyramidal neurons of mouse primary visual cortex (V1)[5]. We made a craniotomy over V1 and infected neurons with adeno-associated virus (AAV2/1-*hSynapsin-1*) (Methods) encoding the jGCaMP8 variants, jGCaMP7f[6] or XCaMP-Gf[20]. After 3 weeks of expression, mice were lightly anaesthetized and mounted under a custom two-photon microscope. Full-field, high-contrast drifting gratings were presented in each of eight directions to the contralateral eye for five

trials (Fig. 3a). Two-photon imaging was performed at frame rates (30 Hz) typical for in vivo imaging (Methods).

Fig. 3: Imaging neural population in the mouse V1 in vivo.



a, Schematic of the experiment (top left). Example image of V1 L2/3 cells (three cells marked by yellow arrows) expressing jGCaMP8s (bottom left), and the same field of view (FOV) colour-coded (three corresponding cells circled) based on the preferred orientation of the neuron (hue) and response amplitude (brightness) (bottom right with colour coding above). This experiment was repeated independently with similar results in 26 FOVs from 6 mice. **b**,**c**, Example traces from three L2/3 neurons expressing jGCaMP8s (**b**; same cells as indicated in **a**) or jGCaMP8f (**c**). Averages of five trials with shaded s.e.m. The polar plots indicate the preferred direction of cells. The orientation selectivity index (OSI) is displayed above each polar plot. **d**, Example zoomed-in fluorescence traces corresponding to the orange boxes in **b** (jGCaMP8s in black) and **c** (jGCaMP8f in blue), normalized to the peak of the response. The inset shows additional detail of the first transient. **e**, Half-decay time of the fluorescence response after the end of the visual stimulus (n = 320 cells from 3 mice (jGCaMP7f), 124 cells from 3 mice (XCaMP-Gf), 317 cells from 5 mice (jGCaMP8f), 365 cells from 3 mice (jGCaMP8m) and 655 cells from 6 mice (jGCaMP8s)).

Kruskal–Wallis multiple-comparison test: $P < 0.001$. Dunn's comparison test: $*P < 0.05$, $***P < 0.001$ and not significant (NS). Full statistics are in the Methods. **f**, Proportion of cells responding to visual stimuli ($n = 12$ FOVs from 3 mice (jGCaMP7f), 19 FOVs from 5 mice (jGCaMP8f), 14 FOVs from 3 mice (jGCaMP8m) and 26 FOVs from 6 mice (jGCaMP8s)). Tukey's multiple-comparison test: $P < 0.001$. One-way ANOVA test was used: $*P < 0.05$, $***P < 0.001$ and NS. Full statistics are in the Methods. **g**, Distribution of response amplitude ($\Delta F/F_0$) for preferred stimulus. The 75th percentile $\Delta F/F_0$ values for each construct: 98% (jGCaMP7f), 38% (jGCaMP8f), 83% (jGCaMP8m) and 183% (jGCaMP8s). $n = 1{,}053$ cells from 3 mice (jGCaMP7f), 1,253 cells from 5 mice (jGCaMP8f), 848 cells from 3 mice (jGCaMP8m) and 1,026 cells from 6 mice (jGCaMP8s). Full statistics are in the Methods.

Visual stimulus-evoked fluorescence transient responses were detected in many cells (Fig. 3b,c) and were stable across trials (Extended Data Fig. 11a). All sensors produced transient responses with rapid rise and decay (Fig. 3b,e). Nearly identical responses were measured after long-term expression of jGCaMP8 (5 additional weeks; Extended Data Fig. 11b–e). XCaMP-Gf was approximately tenfold dimmer than jGCaMP8 or jGCaMP7f (Extended Data Fig. 12a,b), with few responsive cells, whereas protein levels were similar across indicators (Extended Data Fig. 12c,d). These data are consistent with characterization of purified protein (Extended Data Fig. 5) showing that XCaMP-Gf fluorescence is very low. Thus, we did not study XCaMP further.

The contrast changes in visual stimuli were tracked faithfully by fluorescence changes (Fig. 3b,c). Consistent with in vitro characterization, jGCaMP8f showed significantly shorter $t_{1/2,\mathrm{decay}}$ (median of 84 ms, first to third quartile range = 32–153 ms) than jGCaMP7f (median of 110 ms, first to third quartile range = 41–223 ms; $P < 0.05$) and comparable with jGCaMP8m (median of 84 ms, first to third quartile range = 32–165 ms) and XCaMP-Gf (median of 91 ms, first to third quartile range = 48–155 ms; Fig. 3e). jGCaMP8s decay was significantly slower than the other indicators.

We quantified indicator sensitivity to neural activity as the proportion of expressing neurons responsive[5,6] to visual stimuli (Fig. 3f) and as the

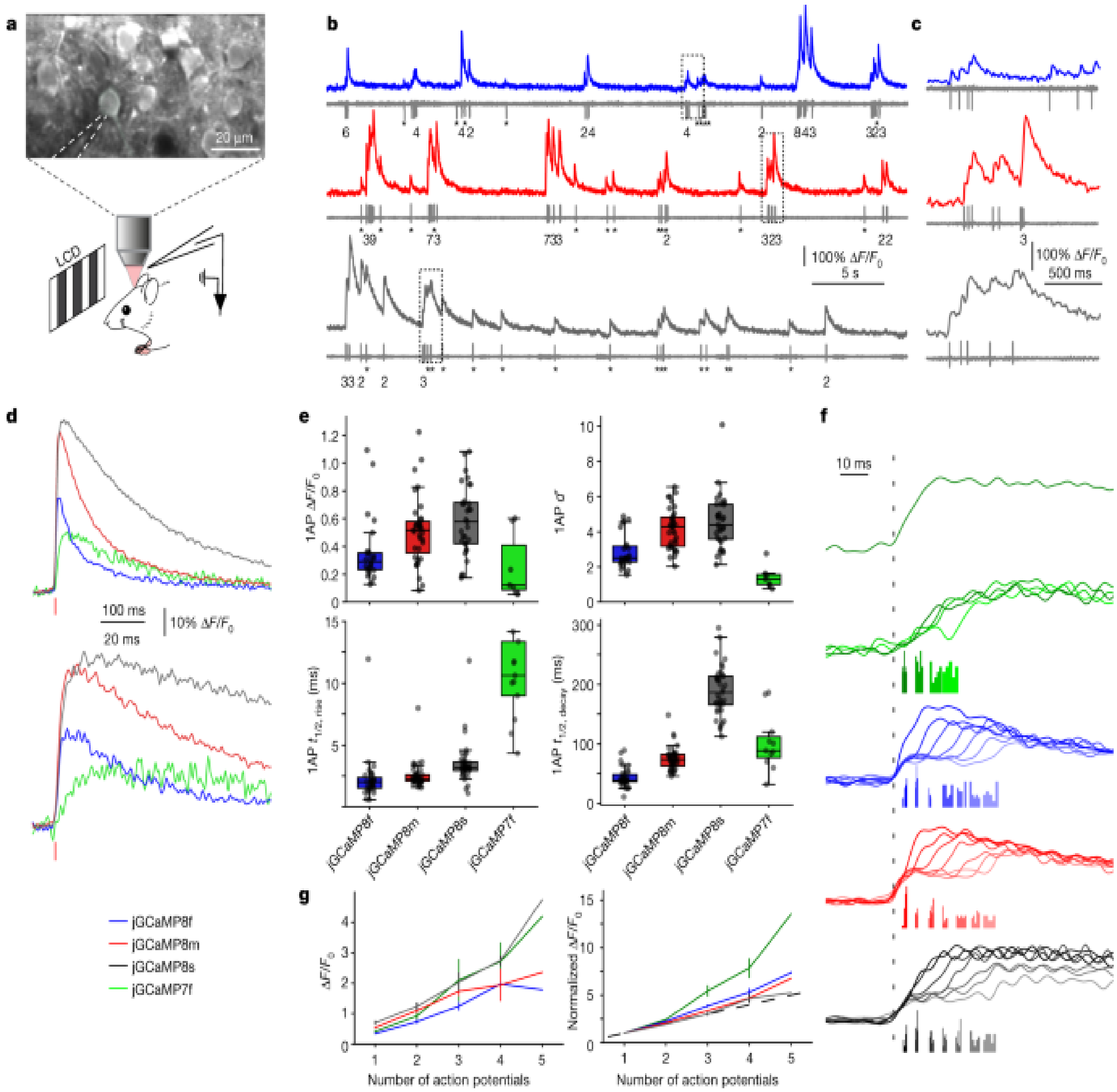
cumulative distribution of peak $\Delta F/F_0$ across cells (Fig. 3g). Significantly more responsive cells were seen for jGCaMP8s and jGCaMP8m than for jGCaMP8f and jGCaMP7f (Fig. 3f; $P < 0.001$). Furthermore, the distribution of peak $\Delta F/F_0$ was shifted towards larger values for jGCaMP8s than for the other indicators (Fig. 3g). Peak amplitude of visually evoked fluorescence transient responses was significantly higher for jGCaMP8s than for other sensors, followed by jGCaMP8m and jGCaMP7f, than by jGCaMP8f (Fig. 3g). The response amplitudes of indicators with short $t_{1/2,\mathrm{decay}}$, particularly jGCaMP8f, were underestimated in these experiments because the relatively slow imaging rate does not reliably catch the peaks of the responses.

Orientation tuning was similar for all sensors, except that jGCaMP8m and jGCaMP8s revealed a larger proportion of neurons with low orientation selectivity (Extended Data Fig. 13). A plausible explanation is that the high-sensitivity indicators detect activity of GABAergic interneurons that is missed by the other sensors. Interneurons yield smaller fluorescence responses[5], and have less sharp orientation tuning than excitatory neurons[30]. This hypothesis is supported by experiments with simultaneous imaging and electrophysiology (see below).

Simultaneous imaging and electrophysiology

To quantify jGCaMP8 responses to neural activity, we combined two-photon imaging (122 Hz) and loose-seal, cell-attached electrophysiological recordings[5] (Fig. 4a). We compared fluorescence changes and spiking across sensors ($n = 40$ cells from 8 mice (jGCaMP8f), 47 cells from 7 mice (jGCaMP8m), 49 cells from 7 mice (jGCaMP8s) and 23 cells from 5 mice (jGCaMP7f); Extended Data Fig. 14a–f and Supplementary Table 2). Fluorescent signals for cell body regions of interest were corrected for neuropil signal[5,6] (Extended Data Fig. 14g–j). All jGCaMP8 variants produced large fluorescence transient responses even in response to single APs (Fig. 4b–d).

Fig. 4: Simultaneous electrophysiology and imaging in the mouse V1 in vivo.



a, Schematic of the experiment. Representative FOV (top) is also shown. The recording pipette is indicated by dashed lines. **b**, Simultaneous fluorescence and spikes, for example, neurons expressing jGCaMP8f (top), jGCaMP8m (middle) and jGCaMP8s (bottom). The number of spikes for each burst is indicated below the trace (single spikes are indicated by asterisks). **c**, Zoomed-in view of traces corresponding to the dashed boxes in **b**. **d**, Grand average of fluorescence response elicited by single APs, aligned to the AP peak (red vertical bar), reconstructed at a temporal resolution of 500 Hz (see text and Extended Data Fig. 15 for details). **e**, Properties of fluorescence responses elicited by single APs. The dots indicate single cells. For the box-and-whisker plots, the box indicates the median and 25–75th

percentile range, and whiskers indicate the shorter of 1.5 times the interquartile range or the extreme data point. $n = 24$ cells from 9 mice (jGCaMP8f), 35 cells from 11 mice (jGCaMP8m), 31 cells from 10 mice (jGCaMP8s) and 11 cells from 3 mice (jGCaMP7f). **f**, Normalized jGCaMP7f response to a single AP (from **d**) (top), and response to AP doublets, binned based on interspike intervals (bottom). Transient responses are normalized and aligned to the first AP of the doublet (dashed line). The timing of the second AP is represented by the histograms below the transient responses. The interspike intervals are selected to be approximately 5, 10, 15, 20, 25, 30 and 35 ms. Responses for jGCaMP7f (green), jGCaMP8f (blue), jGCaMP8m (red) and jGCaMP8s (black) are shown. **g**, Response linearity. Peak response as a function of the number of APs within a 20-ms window (left) is shown. Mean and s.e.m. are displayed. The right graph is the same as the graph on the left, but normalized to 1AP response. $n = 33$, 23, 14, 4 and 2 cells (jGCaMP8f); $n = 41$, 32, 19, 6 and 2 cells (jGCaMP8m); $n = 38$, 34, 18, 3 and 1 cells (jGCaMP8s); and $n = 15$, 13, 6, 4 and 2 cells (jGCaMP7f) for 1, 2, 3, 4 and 5 APs, respectively.

Our experiments allowed us to resolve fluorescence transient responses with much higher effective temporal resolution than the 122-Hz frame rate. Fields of view were arranged so that each individual neuron, including the patched neuron, occupied less than 20% of the scan lines of the frame (Extended Data Fig. 15). As neurons were scanned at random phases with respect to APs, average fluorescence transient responses could be reconstructed at more than 500-Hz effective temporal resolution (Extended Data Fig. 15). All three jGCaMP8 variants had $t_{1/2,\mathrm{rise}} < 5$ ms, more than five times faster than jGCaMP7f under identical conditions (Fig. 4c–e). Peak responses and SNR for all jGCaMP8 indicators were also larger than for jGCaMP7f (Fig. 4d,e). To study spike-time estimation, we first binned AP doublets with respect to their interspike interval, and reconstructed average fluorescence transient responses for spike doublets with 5-ms, 10-ms, 15-ms and 20-ms interspike intervals. The jGCaMP8 indicators resolved individual APs from doublets at spike rates of up to 50 Hz (Fig. 4f). We subsequently grouped spike bursts based on the number of APs (from 1 to 5) in a 20-ms integration window. All sensors showed monotonic increases in fluorescence response with AP count, with the jGCaMP8 sensors responding more linearly than jGCaMP7f

(Fig. 4g). This greater linearity is consistent with neuronal culture and purified protein results.

The *synapsin-1* promoter yields expression in all neurons, including pyramidal cells and fast-spiking (FS; presumably parvalbumin expressing) interneurons, which are interspersed in our imaged regions of interest. Out of our recorded neurons, we identified the subset of FS interneurons by their high spike rates and short spike durations[31] (Extended Data Fig. 16). All three jGCaMP8 sensors produced robust responses (Extended Data Fig. 16b; approximately 3% $\Delta F/F_0$ on average, with responses up to 5%) to single APs in FS interneurons, much larger than GCaMP6s (approximately 1% $\Delta F/F_0$)[5,6].

We also tested the jGCaMP8 variants alongside GCaMP6f and jGCaMP7f in mouse cerebellar Purkinje cell dendritic arbors, where spike-mediated calcium entry occurs over a period of less than 10 ms (ref. [32]) (Extended Data Fig. 17a,b). jGCaMP8m and jGCaMP8f had faster half-decay time than GCaMP6f and jGCaMP7f (Extended Data Fig. 17c,d), and all jGCaMP8 variants showed faster half-rise time than the controls (Extended Data Fig. 17d,e).

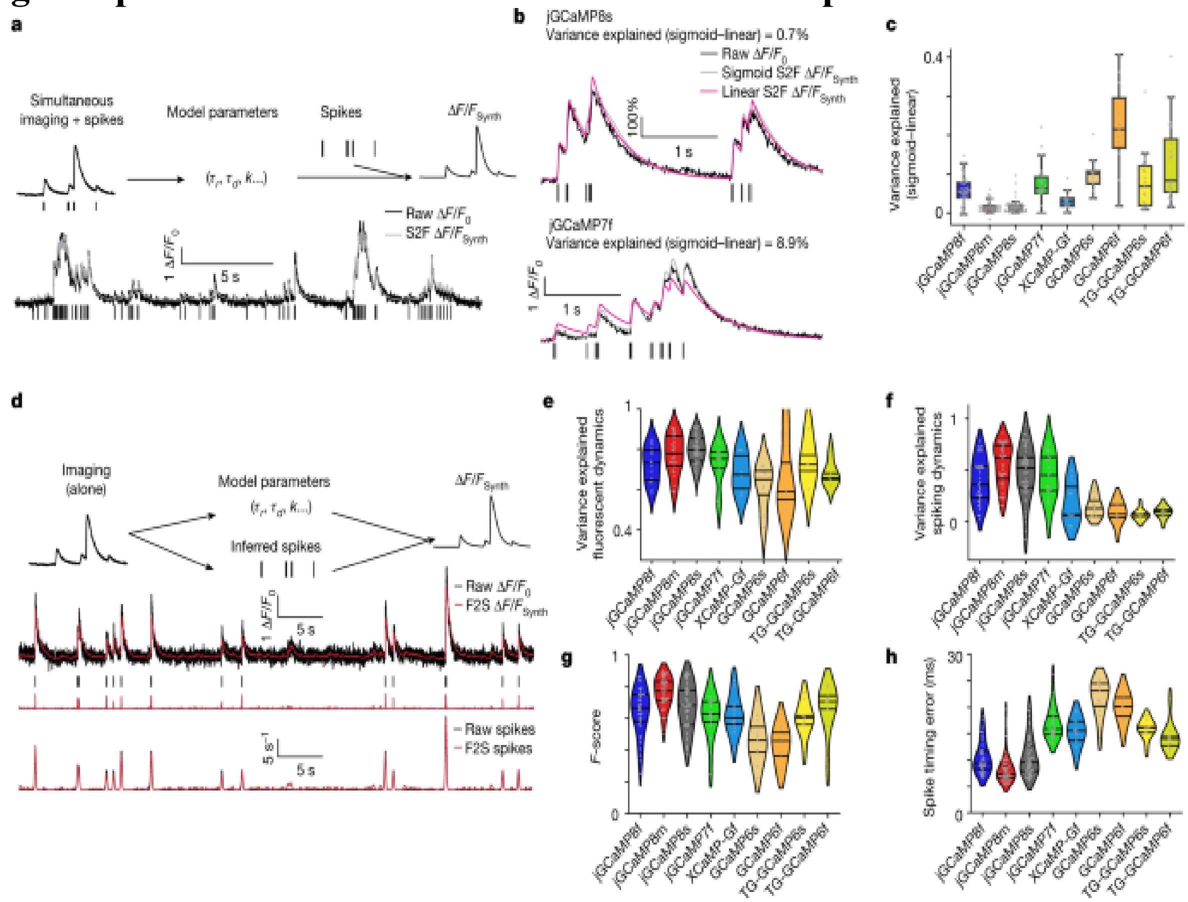
Together, the jGCaMP8 sensors show excellent single-spike detection, spike time estimation, good expression, strong performance in FS interneurons and no evidence of adverse effects of long-term expression.

Spike train modelling with jGCaMP8

Calcium-dependent fluorescence changes are an indirect measure of neural activity[5,33]. A large body of work has been devoted to estimating spike trains from calcium imaging data. Spike extraction is limited by linearity, sensitivity and kinetics of the calcium-dependent sensors[34,35]. We tested the effects of the faster kinetics, superior linearity and higher SNR of the jGCaMP8 indicators on state-of-the-art models of calcium-dependent fluorescence[33] (Methods), using our simultaneous imaging and electrophysiology data (Figs. 4 and 5a). We compared the variance explained across linear and non-linear (sigmoid) models, quantifying to what

extent non-linearities are required to fit fluorescence dynamics for different indicators (Fig. 5b).

Fig. 5: Spike-to-fluorescence and fluorescence-to-spike models.



a, Spike-to-fluorescence (S2F) model. Schematic plot of the S2F forward model that generates a synthetic fluorescence trace ($\Delta F/F_{\text{Synth}}$) from an input spike train (top), and an example fit and data for one cell (bottom) are shown. Measured $\Delta F/F_0$ (black) is overlaid with the simulated $\Delta F/F_{\text{Synth}}$ (grey) from the S2F model. The input to the model, the simultaneously recorded spikes (black), are shown below the traces. **b**, Exemplary cell dynamics with different degrees of non-linearities. **c**, The degree of non-linearity (measured by the difference of variance explained using a sigmoid fit from that using a linear fit). Non-linearity is low for jGCaMP8 sensors (see Extended Data Table 5 for more details) but high for GCaMP6 sensors (TG: GCaMP6 transgenic mouse; otherwise, AAV application). The minima indicate 0th percentile of data (0%); the maxima denote 100%; the centre line indicates 50%; the bounds of box are from 25% (lower quartile)

to 75% (upper quartile); and the whiskers indicate 1.5 times the distance between the upper and lower quartiles. The number of biologically independent cells collected in each condition is shown in Extended Data Table 5. **d**, Fluorescence-to-spike (F2S) model. Schematic plot of the F2S inference model that generates a synthetic fluorescence trace ($\Delta F/F_{\mathrm{Synth}}$) from an inferred spike train (top), and an example fit and data of a cell (bottom) are shown. The first row shows experimental spikes and the measured $\Delta F/F_0$ overlaid with the simulated $\Delta F/F_{\mathrm{Synth}}$ from the F2S model. The second row shows the simultaneously recorded ground-truth spikes (black), shown below the traces, compared with the inferred spikes (red). The third row shows the recorded spike rate overlaid with the inferred spike rate from the F2S model. **e**–**h**, Violin plots, lines from top to bottom: 75%, 50%, 25% of data, respectively. **e**,**f**, Performance of fitting activity using the linear F2S model. Fluorescence dynamics (fits compared with raw fluorescence) (**e**) and spiking (fits compared with ground-truth spiking dynamics) (**f**) are shown. **g**, Performance of spike detectability using the linear F2S model. **h**, Spike-timing error using the linear F2S model.

Linear models performed better for jGCaMP8 than for GCaMP6s or jGCaMP7f in fitting fluorescence traces (Extended Data Table 5), reflecting their linearity (Fig. 5b,c and Extended Data Table 5), SNR and kinetics (Extended Data Table 6 and Extended Data Fig. 18a–g). Model estimates of rise and decay time constants are consistent with direct measurement (Extended Data Fig. 18c,f). Moreover, the model shows that the jGCaMP8 indicators maintain linearity over a wide range of neural activity, in contrast to jGCaMP7f (Fig. 5b,c and Extended Data Fig. 18h–j).

We next examined recovery of spike timing using widely used deconvolution algorithms (Fig. 5d). A linear inference model[34] showed excellent performance in fitting both fluorescence and spiking activity for the jGCaMP8 indicators (Fig. 5e,f and Extended Data Fig. 19a,b). These two measures diverged in some cases, for example, for jGCaMP6s, due to sensor non-linearity (Fig. 5c); this divergence was not reduced by using non-linear inference models (Extended Data Fig. 19f–i). Finally, the jGCaMP8m and jGCaMP8s sensors outperformed the other sensors in spike detection (Fig. 5g and Extended Data Fig. 19c) and timing accuracy (Fig. 5h and Extended Data Fig. 19d).

Discussion

Previous structure–function studies have revealed that the fluorescence kinetics of GCaMP-type indicators is sensitive to mutations at the interface between CaM and the CaM-binding peptide (RS20 in GCaMP; ckkap peptide in XCaMP)[5,8]. For example, the fast variants of GCaMP6f and jGCaMP7f, which were optimized for kinetics, have key beneficial mutations at the CaM–RS20 interface, far from the CaM Ca^{2+}-binding EF hands. These studies suggest that conformational changes at the CaM–RS20 interface constitute a kinetically limiting step between Ca^{2+} binding and fluorescence emission. However, extensive site-saturation mutagenesis of the CaM–RS20 interface failed to dramatically improve kinetics without large sacrifices in SNR[5,6,20]. Inspired by previous work[8], we explored larger sequence changes by replacement of RS20 with 30 diverse CaM-binding peptides. Sensors with a peptide from ENOSP had fast kinetics and good SNR and were further optimized through structure-guided mutagenesis (Fig. 1).

The resulting jGCaMP8 sensors overcome major limitations of previous GECIs. All jGCaMP8 sensors respond to calcium changes with fast kinetics. In vivo fluorescence half-rise times after APs were less than 5 ms (cortical pyramidal neurons; Fig. 4). Such fast kinetics follow neural activity modulations on the rapid timescales relevant to behaviour (Fig. 2). Moreover, the jGCaMP8 sensors are more linear than previous GCaMP sensors (Extended Data Table 3), which facilitates quantitative spike estimation from calcium imaging data (Fig. 5). jGCaMP8 retains many major characteristics of other GCaMP sensors, such as nuclear-excluded expression, bright fluorescence, and excitation and emission spectra. We saw no evidence of cytomorbidity in our experiments, although long-term, high-level expression will probably produce this, as with all GECIs[3,5,36].

jGCaMP8s has the largest single-spike fluorescence change of any calcium indicator, and a moderate half-decay time (200 ms, in mouse brain). The brightness, baseline fluorescence and quantum efficiency of the calcium-bound jGCaMP8 sensors are similar to jGCaMP7 and GCaMP6. Thus, jGCaMP8s sensitivity comes at a cost: saturation at lower spike number and hence lower dynamic range (Extended Data Fig. 3), although this is

ameliorated by their improved linearity and kinetics. We expect jGCaMP8s to become the new standard for most in vivo calcium imaging. jGCaMP8s has an apparent affinity for calcium comparable with resting [Ca^{2+}] in pyramidal neurons in brain slices[15] (46 nM versus 50 nM) (Extended Data Table 3). However, fluorescence changes of several-fold were routinely seen in vivo (Fig. 4), suggesting that resting fluorescence is lower in vivo than with brain slices, or that calcium affinity is weaker than that measured in cuvette.

Compared with jGCaMP8s, jGCaMP8f and jGCaMP8m have faster fluorescence decay and smaller peak fluorescence changes, and higher dynamic ranges. These sensors are ideal to track activity in FS neurons (Extended Data Fig. 16) and applications in which analysis of spike timing is critical[21,37]. Because of their fast fluorescence decay times, the jGCaMP8 indicators will benefit from imaging at higher sampling rates than the widely used jGCaMP7s and GCaMP6s indicators.

Calcium transients are particularly rapid in small structures such as axons, dendrites and spines[18]. The faster jGCaMP8 indicators capture these fleeting signals more efficiently than other sensors. As a result, we observed strong neuropil signals[38] with the jGCaMP8 indicators, which may degrade the SNR in densely labelled neuronal populations[39]. Localizing indicators to the soma, for example, using the RiboL1 tag, may be especially helpful for the jGCaMP8 indicators to optimize the SNR and facilitate segmentation of cell bodies from neuropil[40].

Genetically encoded voltage indicators (GEVIs) can be used to image spikes in single neurons in vivo with fast kinetics. Because calcium is sensed in the three-dimensional cytoplasm, whereas voltage is sensed in the two-dimensional membrane, GECIs have a substantial inherent SNR advantage. Given that the response times of jGCaMP8 sensors approach those of GEVIs[41], with much higher SNR, we believe that population imaging of spiking activity will largely remain the domain of calcium imaging. Voltage imaging will be useful in neurons that lack robust spike-evoked calcium signals[42] and for reporting subthreshold membrane potential changes that are largely invisible to calcium imaging.

In recent years, calcium imaging has become the dominant method to track neural activity, especially in small model systems. However, because of slow GECI kinetics, imaging has been mostly used to map relatively static representations of neural information, rather than tracking the rich dynamics in neural circuits[21,22]. Most imaging is performed on timescales of hundreds of milliseconds, much slower than electrical signalling and information processing in neural circuits. For example, primates can make decisions involving multiple brain areas (including higher cortical areas) in less than 100 ms (ref. [43]), implying that individual neurons process information in milliseconds. So far, neural studies of these fast processes have largely been the domain of electrophysiology. The jGCaMP8 calcium indicators substantially narrow the kinetic gap between imaging and electrophysiology.

Methods

All surgical and experimental procedures were conducted in accordance with protocols approved by the Institutional Animal Care and Use Committee (IACUC) and Institutional Biosafety Committee (IBC) of Janelia Research Campus (all work but the cerebellum), and of the IACUC and IBC at Princeton University (cerebellum work).

Sensor design

We surveyed the Protein Data Bank (PDB) for unique structures of CaM in complex with a single peptide. Twenty-nine peptides were sufficiently different from the RS20 peptide sequence used in previous GCaMPs to warrant testing (Extended Data Table 1). The structures of these complexes were superimposed on the GCaMP2 structure (PDB ID: 3EK4) in PyMOL, and amino acids were added or removed to bring all peptides to a length estimated to work well in the GCaMP topology. Synthetic DNA encoding each of the 29 peptides replaced the RS20 peptide in the bacterial expression vector pRSET-A-GCaMP6s. Of the initial sensors, 20 of the 29 sensed calcium. All 20 had lower saturating fluorescence change than GCaMP6s, all but three had weaker Ca^{2+} affinity (apparent K_d) than GCaMP6s, all but one had lower cooperativity (Hill coefficient (n)), and many were dimmer (Extended Data Table 1). Several sensor variants showed much faster Ca^{2+}

decay kinetics, as determined by stopped-flow fluorescence on purified protein (Extended Data Table 3). On the basis of fast kinetics, saturating fluorescence change, apparent K_d, Hill coefficient and apparent brightness, we prioritized variants based on the peptides from endothelial nitric oxide synthase (PDB ID: 1NIW; peptide 'ENOSP') and death-associated protein kinase 1 (PDB ID: 1YR5; peptide 'DAPKP') for optimization (Extended Data Table 3).

Sensor optimization

These two sensor scaffolds were optimized in protein purified from *Escherichia coli* expression. Libraries were constructed to mutate the linker (linker 1) connecting the peptide to circularly permuted GFP (cpGFP)[4,44] and screened for high signal change and retained fast kinetics. The linker connecting cpGFP and CaM (linker 2) was similarly mutated on top of variants from the optimization of linker 1. Out of 4,000 ENOSP-based variants and 1,600 DAPKP-based variants, 23 and 10, respectively, had fast kinetics and high saturating fluorescence change in purified protein (data not shown).

Guided by the structure of jGCaMP8.410.80, we targeted 16 interface positions for site-saturation mutagenesis: 7 in ENOSP, 4 on cpGFP and 5 on CaM (Extended Data Fig. 1). Sensor variants were tested in cultured neurons for higher sensitivity in detecting neural activity while maintaining fast kinetics. Several single mutations improved properties (Supplementary Table 1), particularly residues near the ENOSP C terminus and the cpGFP–CaM interface. Beneficial point mutations were combined in subsequent rounds of screening. Ten additional CaM positions (Extended Data Fig. 1) surrounding ENOSP were next subjected to site-saturation mutagenesis. Finally, mutations (Extended Data Fig. 1) from the FGCaMP sensor (developed using CaM and RS20-like peptide sequences from the fungus *Aspergillus niger* and the yeast *Komagataella pastoris*)[45,46] were introduced to improve biorthogonality and/or kinetics.

Sensor screen and characterization in solution

Cloning, expression and purification of sensor variants in *E. coli*, calcium titrations, pH titrations, kinetic assay and photophysical analysis were performed as previously described[4,47].

In this study, the RSET tag (His_6 tag-Xpress epitope-enterokinase cleavage site), which had been carried over from the pRSET-A cloning vector in earlier work[3,4,5,48], was removed from all sensors; constructs instead encode a hexa-histidine affinity tag, Met-His_6 tag–peptide–linker 1–cpGFP–linker 2–CaM. For the screen of linkers replacing RS20 (sometimes referred to as ‘M13’), libraries of sensors in the pRSET-A bacterial expression vector were generated using primers containing degenerate codons (NNS) with Q5 site-directed mutagenesis (New England BioLabs) and transformed into T7 Express-competent cells (New England BioLabs). A sequence encoding six repeats of the Gly-Gly-Ser tripeptide was designed as a highly flexible, presumably non-CaM-binding negative control. We expressed the new variants, as well as the presumptive Gly-Gly-Ser negative control and GCaMP6s as a positive control, in *E. coli* T7 Express. Single colonies were picked and grown in 800 µl ZYM-5052 autoinduction medium containing 100 µg ml^{-1} ampicillin in 96-deep-well blocks for 48 h at 30 °C. Cells were collected by centrifugation, frozen, thawed and lysed. Clarified lysate was used to estimate the dynamic range by measuring fluorescence in the presence of 1 mM Ca^{2+} or 1 mM EGTA.

For protein purification, T7 Express cells containing sensors were grown at 30 °C for 48 h in ZYM-5052 autoinduction medium with 100 µg ml^{-1} ampicillin. Collected cells were lysed in 1/50 volume of B-PER (Thermo Fisher) with 1 mg ml^{-1} lysozyme and 20 U ml^{-1} Pierce Universal Nuclease (Thermo Fisher) and subsequently centrifuged. Supernatants were applied to HisPur Cobalt Resin (Thermo Fisher). The resin was washed with 20 column volumes of 20 mM Tris, pH 8.0, 300 mM NaCl and 1 mM imidazole, followed by 10 column volumes of 20 mM Tris, pH 8.0, 500 mM NaCl and 5 mM imidazole. Proteins were eluted into 20 mM Tris, pH 8.0, 100 mM NaCl and 100 mM imidazole.

For calcium titrations, sensors were diluted 1:100 in duplicate into 30 mM MOPS, pH 7.2, 100 mM KCl containing either 10 mM CaEGTA (39 µM free calcium) or 10 mM EGTA (0 µM free calcium). As before, these two

solutions were mixed in different amounts to give 11 different free calcium concentrations. GCaMP fluorescence (485-nm excitation, 5-nm bandpass; 510-nm emission, 5-nm bandpass) was measured in a Tecan Safire2 plate reader (Tecan). The data were fit with a sigmoidal function using KaleidaGraph (Synergy Software) to extract the apparent K_d for Ca^{2+}, the Hill coefficient and the saturating fluorescence change.

The half-decay time of fluorescence after calcium removal ($t_{1/2,decay}$) was determined at room temperature using a stopped-flow device coupled to a fluorimeter (Applied Photophysics). Each sensor variant in 1 μM Ca^{2+} in 30 mM MOPS, pH 7.2, and 100 mM KCl was rapidly mixed with 10 mM EGTA in 30 mM MOPS, pH 7.2, and 100 mM KCl. Fluorescence decay data were fit with a single or double exponential decay function.

For pH titrations, purified proteins were diluted into pH buffers containing 50 mM citrate, 50 mM Tris, 50 mM glycine, 100 mM NaCl and either 2 mM $CaCl_2$ or 2 mM EGTA, which were pre-adjusted to 24 different pH values between 4.5 and 10.5 with NaOH. A sigmoidal function was used to fit fluorescence versus pH, and the pK_a value was determined from the midpoint.

Sequence and structural analysis of variants

Linker1 encodes Leu-Glu in GCaMP6s- and indeed, in all previous RS20-based GCaMP sensors. This linker was extensively mutated in the GCaMP5 screen[4] but the best variant, GCaMP5G, retained Leu-Glu; we first mutated Leu-Glu to fully degenerate 2-amino acid (aa) sequences and screened for variants with both high signal change and retained fast kinetics. Following selection of the best 2-aa linkers, these variants were expanded to libraries of 3-aa linkers by addition of fully degenerate codons. After optimization of linker 1, linker 2 was mutated from Leu-Pro, to which it had been selected in GCaMP5G[28], the parent of GCaMP6 and GCaMP7. Mutagenesis of linker 2 was similar to that for linker 1, but alternative linker 2 sequences either slowed kinetics or decreased $\Delta F/F_0$, and linker 2 was thus retained as Leu-Pro.

In addition to jGCaMP8f, jGCaMP8m and jGCaMP8s, several other variants may be of interest, including 455, 543, 640, 707 and 712 (Supplementary Table 1). All promising variants contain, in addition to the Leu-Lys-Ile linker 1, additional mutations to the ENOSP peptide: Asn19Thr and Ser24Ile appear in every variant except 712, Ser26Arg appears in every variant but jGCaMP8s (with Ser26Met), jGCaMP8m has Ala25Gly, and 712 has Met28Ser. Every variant contains the Gln88Glu mutation at the CaM–GFP interface. Further mutations include Phe286Tyr (jGCaMP8s, jGCaMP8m and 707); Glu288Gln (707); Gln315Leu (jGCaMP8f), Gln315His (jGCaMP8s and 707), Gln315Lys (455); Met346Gln (543); and Met419Ser (640). Of these, Phe286Tyr comes from the FGCaMP sensor; all others are unique to this work. GCaMP6s data from both purified protein and cultured neurons are essentially identical between this work (lacking the RSET tag) and previous work (with the RSET tag) (data not shown), implying that the RSET tag does not noticeably modulate GCaMP function in protein and neuronal culture and that observed jGCaMP8 improvements stem from the peptide substitution and other mutations.

Photophysical measurements

All measurements were performed in 39 μM free calcium (+Ca) buffer (30 mM MOPS, 10 mM CaEGTA in 100 mM KCl, pH 7.2) or 0 μM free calcium (−Ca) buffer (30 mM MOPS, 10 mM EGTA in 100 mM KCl, pH 7.2). Absorbance measurements were performed using a UV–Vis spectrometer (Cary 100, Agilent technologies), and fluorescence excitation–emission spectra were measured using a spectrofluorometer (Cary Eclipse, Varian). $\Delta F/F_0$ was calculated from the fluorescence emission spectra of the proteins in +Ca and −Ca buffer. Quantum yield for +Ca solutions was measured using a spectrometer with an integrating sphere (Quantaurus, Hamamatsu); for −Ca, a relative method was applied using jGCaMP7f (quantum yield of 0.60) as a reference. Extinction coefficients were determined via the alkali denaturation method, using the extinction coefficient of denatured eGFP as a reference ($\varepsilon = 44{,}000\ M^{-1}\ cm^{-1}$ at 447 nm).

Photobleaching measurements

Solutions of 2–4 µM protein were prepared in 30 mM MOPS, pH 7.2 (+Ca and –Ca buffer) with 0.1% BSA added. A sample of these solutions was added to 1-octanol in a 1:9 ratio, and this mixture was vortexed briefly. The resulting emulsion was placed on a pre-silanized glass microscopy slide and fitted with a coverslip. Fluorophore bleaching was accomplished by illuminating a microdroplet using an upright microscope (Zeiss Axio Observer Z2) and a ×20, NA 0.8 objective. Light illumination was obtained using a 488-nm (Sapphire 488, Coherent) laser excitation at 11.45-mW power (intensity = 14.0 W cm^{-2}), and the emission was collected using a 525BP50 filter. Fluorescence was detected by a fibre-coupled avalanche photodiode (SPQM-AQRH14, Pacer). Obtained bleaching profiles were fit using a double-exponential fitting function in MATLAB to calculate their respective time constants (τ_{bleach}).

For quantitative comparison of photobleaching for fluorophores, the bleaching probability (P_b) needs to be computed as referenced in ref. [49]. To quantify P_b for each fluorophore, their respective excitation rate (W) was obtained. The excitation rate W for a laser source can be calculated as a product of extinction coefficient (ε (λ) in M^{-1} cm^{-1}) and intensity (I (λ) in W cm^{-2}) for the discrete excitation wavelength (λ, in nm) as shown in the equation (1):

$$W=0\,.\,0192\phantom{\rule{ 0.25em}{0ex}}\varepsilon \,I\,\lambda $$
(1)

Photobleaching is further characterized by the number of photons (N_p) emitted before photobleaching, which is the product of fluorescence quantum yield (ϕ_f), excitation rate (W) and photobleaching time constant (τ_{bleach}) as shown in the equation (2):

$${N}_{p}={\varphi }_{f}\,W\,{\tau }_{{\rm{bleach}}}$$
(2)

In a rate equation model for bleaching proceeding from singlet or triplet states, the photobleaching probability P_b is inversely related to the total

number of fluorescent photons emitted by $N_{p}=\varphi_{f}/P_{b}$. Using equation (2), this can be expressed as:

$$P_{b}=1/W\,\tau_{\rm{bleach}} \tag{3}$$

The photobleaching probability P_b is the most rigorous, as it is independent of the fluorescence quantum yield. The calculated P_b values for the proteins are presented in Extended Data Table 3.

Two-photon spectroscopy

The two-photon excitation spectra were performed as previously described[1]. Protein solutions of 1–5 μM concentration in +Ca or −Ca buffer were prepared and measured using an inverted microscope (IX81, Olympus) equipped with a ×60, 1.2 NA water immersion objective (Olympus). Two-photon excitation was obtained using an 80 MHz Ti-Sapphire laser (Chameleon Ultra II, Coherent) with sufficient power from 710 nm to 1,080 nm. Fluorescence collected by the objective was passed through a short-pass filter (720SP, Semrock) and a bandpass filter (550BP200, Semrock) and detected by a fibre-coupled avalanche photodiode (SPCM_AQRH-14, Perkin Elmer). The obtained two-photon excitation spectra were normalized to 1 μM concentration and subsequently used to obtain the action cross-section spectra with fluorescein as a reference (average action cross-section spectra from refs. [50,51]).

Fluorescence correlation spectroscopy was used to obtain the two-photon molecular brightness of the protein molecule. The peak molecular brightness was defined by the rate of fluorescence obtained per total number of emitting molecules. Protein solutions (50–100 nM) were prepared in +Ca buffer and excited with 930-nm light at powers ranging from 2 mW to 30 mW for 200 s. The obtained fluorescence emission was collected by an avalanche photodiode and fed to an autocorrelator (Flex03LQ, Correlator.com). The obtained autocorrelation curve was fit to a diffusion model through a built-in MATLAB function[52] to determine the number of molecules <*n*> present in the focal volume. The two-photon molecular brightness (ε) at each laser power was calculated as the average rate of

fluorescence $\langle F \rangle$ per emitting molecule $\langle n \rangle$, defined as $\varepsilon = \langle F \rangle / \langle n \rangle$ in kilocounts per second per molecule. As a function of laser power, the molecular brightness initially increases with increasing laser power, then levels off and decreases due to photobleaching or saturation of the protein chromophore in the excitation volume. The maximum or peak brightness achieved, $\langle \varepsilon_{max} \rangle$, represents a proxy for the photostability of a fluorophore.

Screening in neuronal cell culture

GCaMP variants were cloned into an hSyn1-GCaMP-NLS-mCherry-WPRE expression vector, and XCaMP variants (XCaMP-G, XCaMP-Gf and XCaMP-Gf0) were cloned into the same expression vector with the nuclear export sequence that was attached to the XCaMP sensors in the original publication[8]. As this excludes the XCaMP sensors from the nucleus, where Ca^{2+} signals are slower[53], whereas the variants developed here were not explicitly excluded (although GCaMPs without an explicit nuclear export sequence are nevertheless fairly nuclearly excluded), this will make the XCaMPs appear faster than they really are than the GCaMP indicators.

The primary rat culture procedure was performed as previously described[6]. In brief, neonatal rat pups (Charles River Laboratory) were euthanized, and neocortices were dissociated and processed to form a cell pellet. Cells were resuspended and transfected by combining 5×10^5 viable cells with 400 ng plasmid DNA and nucleofection solution in a 25-µl electroporation cuvette (Lonza). Electroporation of GCaMP mutants was performed according to the manufacturer’s protocol.

Neurons were plated onto poly-d-lysine (PDL)-coated, 96-well, glass-bottom plates (MatTek) at approximately 1×10^5 cells per well in 100 µl of a 4:1 mixture of NbActiv4 (BrainBits) and plating medium (28 mM glucose, 2.4 mM $NaHCO_3$, 100 $\mu g\ ml^{-1}$ transferrin, 25 $\mu g\ ml^{-1}$ insulin, 2 mM l-glutamine, 100 $U\ ml^{-1}$ penicillin, 10 $\mu g\ ml^{-1}$ streptomycin and 10% FBS in MEM). Typically, each plate included GCaMP6s (eight wells), GCaMP6f (eight wells) and jGCaMP7f (eight wells). Other wells were electroporated with mutated variants (four wells per variant), for a total of 80 wells (the first and last columns in the plate were not used to reduce edge effects).

Plates were left in the incubator at 37 °C and 5% CO_2. The next day, 190 µl of NbActiv4 medium was added to each well.

On 12–15 days in vitro (DIV), neurons were rinsed three times with imaging buffer containing 140 mM NaCl, 0.2 mM KCl, 10 mM HEPES and 30 mM glucose (pH 7.3–7.4) and left in a solution containing imaging buffer with added receptor blockers (10 µM CNQX, 10 µM (R)-CPP, 10 µM gabazine and 1 mM (S)-MCPG; Tocris) to reduce spontaneous activity; neurons then underwent field stimulation and imaging[6,24]. Fluorescence timelapse images (200 Hz; total of 7 s) were collected on an Olympus IX81 microscope using a ×10, 0.4 NA objective (UPlanSApo, Olympus) and an ET-GFP filter cube (#49002, Chroma). A 470-nm LED (Cairn Research) was used for excitation (intensity at the image plane of 0.34 mW mm^{-2}). Images were collected using an EMCCD camera (Ixon Ultra DU897, Andor) with 4 × 4 binning, corresponding to a 0.8 mm × 0.8 mm FOV. Reference images (100-ms exposure) were used to perform segmentation. Red illumination for variants co-expressing mCherry was performed with a 590-nm LED (Cairn Research) through an ET-mCherry filter cube (#49008, Chroma) with an intensity of 0.03 mW mm^{-2}. Trains of 1, 3, 10 and 160 field stimuli were delivered with a custom stimulation electrode. For sensor linearity measurements, 1, 2, 3, 5, 10 and 40 field stimuli were delivered. All measurements were performed at room temperature, which contributed to slower kinetics than that reported in vivo (34 °C).

The responses of individual variants were analysed as previously described[5,6]. The Ilastik toolkit[54] was used to segment cell bodies in the reference images. Wells with fewer than five detected neurons, and wells with poor neuronal proliferation, were discarded (labelled as ‘failed segmentation’ in Supplementary Table 1). Plates with more than four failed control (GCaMP6s) wells were discarded and re-screened.

When calculating $\Delta F/F_0$ (defined as $(F_{peak} - F_0)/F_0$), F_{peak} was taken from the single frame with the highest fluorescence intensity post-stimulus, and F_0 was the average intensity of ten frames preceding the stimulus. d' was calculated as follows:

$${d}^{{\prime} }=\frac{{\bar{F}}_{{\rm{top}}}-{\bar{F}}_{{\rm{bottom}}}}{\sqrt{\frac{1}{2}({\sigma }^{2}({F}_{{\rm{top}}})+{\sigma }^{2}({F}_{{\rm{bottom}}}))}}$$

where F_{top} and F_{bottom} are peak and baseline fluorescence traces, respectively, six frames in duration.

As the fluorescent signal was sampled at 200 Hz, fast rise times (less than 10 ms) could not be reliably computed for single trials. Thus, to compute half-rise time ($t_{1/2,rise}$), we found the two frames having fluorescence intensities below and above $F_{peak}/2$, linearly interpolated the trace between them, and computed the timepoint at which the fluorescence would have crossed the $F_{peak}/2$ threshold. Using this technique and averaging across hundreds of neurons for each variant allowed us to approximate ($t_{1/2,rise}$) with higher resolution than the sampling interval.

Median values from each well were reported to quantify performance. Each observation was normalized to the median GCaMP6s value from the same experimental batch. Baseline brightness for constructs co-expressing mCherry was calculated by dividing the GFP cellular fluorescence in the beginning of the 3AP stimulation epoch by the mCherry cellular fluorescence (for a ratiometric measurement). For comparison with XCaMP variants (Extended Data Fig. 4a), no mCherry normalization was performed, but all baseline brightness values were still normalized to GCaMP6s in the same transfection week. To determine significant differences in observations between constructs, a two-tailed Mann–Whitney *U*-test was performed between constructs and controls (GCaMP6s or jGCaMP7f). A median $\Delta F/F_0$ trace was computed across all detected cell bodies in a well for each stimulus. Photobleaching was corrected in the 1AP recordings by fitting a double exponential to the beginning and end segments of the fluorescence trace.

Finally, data were filtered according to three criteria to remove variants without detectable response to 1AP. We filtered out variants (1) with $t_{1/2,rise} < 0.1\times$ or $> 4\times$ of GCaMP6s, (2) with time-to-peak $>3\times$ of GCaMP6s, and (3) with $t_{1/2,decay} < 0.01\times$ of GCaMP6s, as these represented inaccurate fits to

non-responsive fluorescence traces (labelled as ‘no detectable response’ in Supplementary Table 1).

To evaluate the baseline fluorescence of the jGCaMP8 series compared with jGCaMP7f and the XCaMP series, all constructs were transfected side-by-side (2 consecutive transfection weeks, five 96-well plates). To minimize possible plate-to-plate variability within each transfected batch, the baseline fluorescence of each construct was normalized to in-plate GCaMP6s.

We have implemented several improvements to the neuronal culture screening rig over the years (Supplementary Table 4). These improvements to the rig result in slight changes in values of $\Delta F/F_0$ and other parameters for our control GECIs (for example, GCaMP6 and jGCaMP7) compared with in our original publications. Note that at all times, we compared variants to reference sensors in an apples-to-apples comparison, using data obtained from in-plate controls.

All of the parameters measured in our screen can be examined as an interactive scatterplot in a Binder notebook (https://mybinder.org/v2/gh/ilyakolb/jGCaMP8-neuron-culture-screen/HEAD?labpath=interactive-multiparameter-screening-plot.ipynb). The data are also collated in Microsoft Excel in Supplementary Table 1.

Fluorescence recovery after photobleaching

Fluorescence recovery after photobleaching experiments were carried out on a Nikon Ti-E inverted microscope outfitted with a Yokogowa CSU-X1 spinning disk and an Andor DU-897 EMCCD camera. Fluorescence excitation was carried out using a solid-state laser line at 488 nm, and emission was collected with a ×100 1.49 NA objective (Nikon Instruments) through a standard GFP filter set. Photobleaching was performed using a Bruker Mini-Scanner by focusing a 405-nm laser to a single, diffraction-limited spot for 100 ms. Cultured neurons plated in 35-mm glass-bottom dishes (MatTek) were immersed in regular imaging buffer with the addition of synaptic blockers (same as used for neuronal culture field stimulation) and 1 µM TTX to block AP generation. In a subset of experiments, the buffer was supplemented with 5 µM ionomycin. Bleaching spots were chosen to be on the soma of the neuron but distant from the nucleus. A spot

was photobleached ten times (0.1 Hz) as the cell was concurrently imaged at 25 or 50 frames per second.

For analysis, pixels within a 1.5-µm radius around the bleach spot were averaged in each frame. The resulting fluorescence trace was normalized to the mean fluorescence of an identically sized spot on the opposite side of the soma, outside the nucleus. The trace was then split into ten epochs (each corresponding to a bleaching event) and the fluorescence $f_i(t)$ of each epoch i was normalized by dividing by the fluorescence value immediately preceding the bleaching pulse (${f}_{i}\left({t}_{{\rm{pre}}}\right)$) as follows:

$${\bar{f}}_{i}(t)=\frac{{f}_{i}(t)}{{f}_{i}({t}_{{\rm{pre}}})}$$

The resistant fraction (RF) was calculated as follows:

$${\rm{R}}{\rm{F}}\,({\rm{ \% }})=100\left(1-{\bar{f}}_{1}({t}_{\text{fin}})-\frac{1}{9}\mathop{\sum }\limits_{i=2}^{10}(1-{\bar{f}}_{i}({t}_{\text{fin}}))\right)$$

where ${\bar{f}}_{i}({t}_{\text{fin}})$ is the final fluorescence value at the end of epoch i, and the final term in the equation is the averaged fluorescence loss of all epochs after the first. This term is subtracted to account for the overall fluorescence loss with each bleaching pulse.

Crystal structure determination

All GCaMP samples for crystallization were kept in 20 mM Tris, 150 mM NaCl, pH 8.0, and 2 mM $CaCl_2$. All crystallization trials were carried out at 22 °C with the hanging-drop vapour diffusion method. Commercial sparse-matrix screening solutions (Hampton Research) were used in initial screens. Of the protein solution, 1 µl was mixed with 1 µl of reservoir solution and equilibrated against 250 µl of reservoir solution. Diffraction data were collected at beamline 8.2.1 at the Berkeley Center for Structural Biology and processed with XDS[55]. The phase was determined by molecular replacement using MOLREP, and the structure of GCaMP2 (PDB ID: 3EK4)[56] without the RS20 peptide as the starting model. Refinement was performed using

REFMAC[57] followed by manual remodelling with Coot[58]. Details of the crystallographic analysis and statistics are presented in Extended Data Table 2. The crystal structure has been released on the PDB website (rcsb.org), entry 7ST4. wwPDB validation scores are excellent (https://files.rcsb.org/pub/pdb/validation_reports/st/7st4/7st4_full_validation.pdf).

Adult *Drosophila* L2 assay

GECIs were tested by crossing males carrying the variant to a w+;*53G02*-Gal4^{AD} (in attP40);*29G11*-Gal4^{DBD} (in attP2) females[59]. Heterozygous flies were used in our experiments. Sensor cDNAs were codon-optimized for *Drosophila*. Flies were raised at 21 °C on standard cornmeal molasses medium.

Three to five days after eclosure, females were anaesthetized on ice. After transferring to a thermoelectric plate (4 °C), legs were removed, and then facing down, the head was glued into a custom-made pyramid using UV-cured glue. The proboscis was pressed in and fixed using UV-cured glue. After adding saline (103 mM NaCl, 3 mM KCl, 1 mM NaH_2PO_4, 5 mM TES, 26 mM $NaHCO_3$, 4 mM $MgCl_2$, 2.5 mM $CaCl_2$, 10 mM trehalose and 10 mM glucose, pH 7.4, 270–275 mOsm) to the posterior side of the head, the cuticle was cut away above the right side, creating a window above the target neurons. Tracheae and fat were removed, and muscles M1 and M6 were cut to minimize head movement.

Two-photon imaging took place under a ×40 0.8 NA water-immersion objective (Olympus) on a laser-scanning microscope (BrukerNano) with GaAsP photomultiplier tubes. Laser power at 920 nm was kept constant at 8 mW using a Pockels cell. No bleaching was evident at this laser intensity. The emission dichroic was 580 nm and emission filters 511/20–25 nm. Images were 32 × 128 pixels with a frame rate at 372 Hz.

A MATLAB script drove the visual stimulation via a digital micromirror device (DMD, LightCrafter) at 0.125 Hz onto a screen covering the visual field in front of the right eye. A blue LED (M470L3, Thorlabs) emitting

through a 474/23–25-nm bandpass filter (to keep blue light from contaminating the green imaging channel) provided illumination.

Light dimming produced a stereotypical calcium increase in L2 neurons[27]. Intensity measurements were taken in medulla layer 2 (Fig. 2a). A target region image was chosen by testing each focal layer with 0.5-Hz full-field visual stimulation until a layer with maximum $\Delta F/F_0$ was identified. Then, 2–3 columns producing a maximum response were identified within this layer. In addition to the region of interest (ROI) containing these L2 columns, a background ROI was selected where no fluorescence was evident. The mean background intensity was subtracted from the mean L2 ROI. Imaging then targeted this region over a protocol involving multiple tests, as shown in Supplementary Table 3.

Image analysis was performed using custom Python scripts. In the $\Delta F/F_0$ calculation, baseline F_0 included the last 20% of images taken at the end of the light period. Stimulus onset is the light-to-dark transition. Change in fluorescence ΔF is the intensity minus baseline. $\Delta F/F_0$ is ΔF divided by baseline. The final signal is processed through a Gaussian filter ($\sigma = 3$). Discriminability index (d') values were calculated the same as in mouse imaging (see below).

Imaging in the *Drosophila* larval neuromuscular junction

We made 20XUAS-IVS-Syn21-op1-GECI-p10 in VK00005 transgenic flies[60] and crossed them with 10XUAS-IVS-myr::tdTomato in su(Hw)attP8 × *R57C10*-Gal4 at VK00020; *R57C10*-Gal4 at VK00040 double-insertion pan-neuronal driver line. Heterozygous flies were used in our experiments. Sensor cDNAs were codon-optimized for *Drosophila*. The neuromuscular junction (NMJ) assay is as in our previous study[10]. In brief, female third instar larvae were dissected in chilled (4 °C) Schneider’s insect medium (Sigma) to fully expose the body wall muscles. Segment nerves were severed in proximity to the ventral nerve cord. Dissection medium was then replaced with room temperature HL-6 saline in which 2 mM $CaCl_2$ and 7 mM l-glutamate were added to induce tetany, freezing the muscles in place. A mercury lamp (X-CITE exacte) light source was used for excitation, and

out-of-objective power was kept less than 5 mW to reduce bleaching. Type Ib boutons on muscle 13 from segment A3–A5 were imaged while the corresponding hemi-segment nerve was stimulated with square voltage pulses (4 V, pulse width of 0.3 ms, duration of 2 s and frequency of 1–160 Hz) through a suction electrode driven by a customized stimulator. Bath temperature and pH were continuously monitored with a thermometer and pH metre, respectively, and recorded throughout the experiment. The filters for imaging were as follows: excitation of centre wavelength (CWL) = 472 nm and bandwidth (BW) = 30 nm, dichroic of 495 nm, and emission of CWL = 520 nm and BW = 35 nm. Images were captured with an EMCCD (Andor iXon 897) at 128.5 frames per second and acquired with Metamorph software. ROIs around boutons were manually drawn, and data were analysed with a custom Python script. Discriminability index values were calculated the same as in mouse imaging.

NMJ immunofluorescence

Variants were crossed to a pan-neuronal driver line, also containing tdTomato (pJFRC22-10XUAS-IVS-myr::tdTomato in su(Hw)attP8;; *R57C10* at VK00020, *R57C10* at VK00040). Third instar larvae were filleted and fixed following standard techniques[61]. Primary chicken anti-GFP (1:1,000; A10262, Thermo Fisher) and secondary goat anti-chicken AlexaFluor Plus 488 (1:800; A32931, Thermo Fisher) were used to stain GECIs. Primary rabbit anti-RFP (1:1,000; 632496, Clontech) and secondary goat anti-rabbit Cy3 (1:1,000; 111-165-144, Jackson) labelled tdTomato.

MBON-γ2α′1 immunofluorescence

Variants were co-expressed with membrane-localized myr::tdTomato using the *MB077B* driver. Adults 3–6 days old were harvested, brains dissected and fixed using standard techniques. GCaMP variants were directly labelled with rabbit anti-GFP (1:500; AlexaFluor 488; A-21311, Molecular Probes). Primary rat anti-RFP (1:500; mAb 5F8, Chromotek) and secondary goat anti-rat Cy3 (1:1,000; 112-165-167, Jackson) labelled tdTomato.

Immunofluorescence quantification

ROIs were drawn on targeted regions using custom Python scripts. Within each ROI, otsu-thresholding was used to identify regions expressing myr::tdTomato. Intensity measurements were then taken for both the variant and tdTomato within these regions. The ratio is the intensity from the green channel (variant staining) divided by the intensity from the red channel (myr::tdTomato staining).

Western blot

Protein was extracted from female brains with the same genotype used in the NMJ immunostaining. Western blots were performed following standard techniques. Each variant was stained using primary rabbit anti-GFP (PC408, Millipore Sigma) and secondary goat anti-rabbit IgG conjugated to horseradish peroxidase (HRP; 31460, Thermo Fisher/Invitrogen). Actin was stained using mouse IgM anti-α-actin (1:5,000; MA1-744, Thermo Fisher/Invitrogen) and goat anti-mouse IgG and IgM-HRP (1:5,000; 31430 and 62-6820, respectively, Thermo Fisher/Invitrogen). Signal was formed using SuperSignal West Dura luminescence and was imaged on a Bio-Rad Gel imager. Band intensity was measured using Fiji[62]. Band intensity from the variant was divided by band intensity from the actin band to determine the ratio.

Mouse work

All mice were cared for in compliance with the Guide for the Care and Use of Laboratory Animals. All experiments at Janelia were approved by the Janelia Research Campus IACUC and IBC committees. Janelia is an AAALAC-accredited institution. Mice were maintained under specific pathogen-free conditions. Mice were housed on a free-standing, individually ventilated (approximately 60 air changes hourly) rack (Allentown). The holding room was ventilated with 100% outside filtered air with 15–20 air changes hourly. Each ventilated cage (Allentown) was provided with corncob bedding (Shepard Specialty Papers), at least 8 g of nesting material (Bed-r'Nest, The Andersons) and red Mouse Tunnel (Bio-Serv). Mice were maintained on a 12:12-h light:dark cycle. The holding room temperature was maintained at 68–72 °F with a relative humidity of 30–70%. Irradiated rodent laboratory chow (LabDiet 5053) was provided ad libitum.

At Princeton, experimental procedures were approved by the Princeton University IACUC (protocol 3080-16) and performed in accordance with the animal welfare guidelines of the US National Institutes of Health. All mice were housed under a 12:12-h regular light cycle for breeding and transferred to reverse light-cycle conditions with 12:12-h reverse light:dark cycle facility, and experiments were performed during the dark cycle. At least 1 week before the experimental days, mice were housed in darkness in an enrichment box containing bedding, houses and wheels (Bio-Serv Fast-Trac K3250/K3251). At other times, mice were housed in cages in the animal facility in groups of 2–4 mice per cage. Mice were maintained at 68–72 °F with a relative humidity of 30–70%.

Mouse surgeries for cortical imaging

Young adult (postnatal day 50–214) male C57BL/6J (Jackson Labs) mice were anaesthetized using isoflurane (2.5% for induction and 1.5% during surgery). A circular craniotomy (diameter of 3 mm) was made above V1 (centred 2.5 mm left and 0.5 mm anterior to the Lambda suture). Viral suspension (30 nl) was injected in 4–5 locations on a 500-µm grid, 300–400 µm deep. Constructs included: AAV2/1-*hSynapsin-1*-jGCaMP8 constructs (pGP-AAV-syn1-jGCaMP8f-WPRE, Addgene plasmid #162376, 4×10^{12} GC per millilitre titre; pGP-AAV-syn1-jGCaMP8m-WPRE, Addgene plasmid #162375, 2.2×10^{12} GC per millilitre titre; pGP-AAV-syn1-jGCaMP8s-WPRE, Addgene plasmid #162374, 2.1×10^{12} GC per millilitre titre). A 3-mm diameter circular coverslip glued to a donut-shaped 3.5-mm diameter coverslip (no. 1 thickness, Warner Instruments) was cemented to the craniotomy using black dental cement (Contemporary Ortho-Jet). A custom titanium head post was cemented to the skull. An additional surgery was performed for loose-seal recordings. Eighteen to eighty days after the virus injection, the mouse was anaesthetized with a mixture of ketamine–xylazine (0.1 mg ketamine and 0.008 mg xylazine per gram body weight), and we surgically removed the cranial window and performed durotomy[63]. The craniotomy was filled with 10–15 µl of 1.5% agarose, then a D-shaped coverslip was secured on top to suppress brain motion and leave access to the brain on the lateral side of the craniotomy.

Two-photon imaging in mouse cortex

Mice were kept on a warm blanket (37 °C) and anaesthetized using 0.5% isoflurane and sedated with chlorprothixene (20–30 μl at 0.33 mg ml^{-1}, intramuscular). Imaging was performed with a custom-built two-photon microscope with a resonant scanner. The light source was an Insight femtosecond-pulse laser (Spectra-Physics) running at 940 nm. The objective was a ×16 water immersion lens with 0.8 numerical aperture (Nikon). The detection path consisted of a custom filter set (525/50 nm (functional channel), 600/60 nm (cell-targeting channel) and a 565-nm dichroic mirror) ending in a pair of GaAsP photomultiplier tubes (Hamamatsu). Images were acquired using ScanImage (vidriotechnologies.com)[64]. Functional images (512 × 512 pixels, 215 × 215 μm^2; or 512 × 128 pixels, 215 × 55 μm^2) of L2/3 cells (50–250 μm under the pia mater) were collected at 30 Hz or 122 Hz. Laser power was up to 50 mW at the front aperture of the objective unless stated otherwise for the XCaMP-Gf experiments.

Loose-seal recordings in mouse cortex

Micropipettes (3–9 MΩ) were filled with sterile saline containing 20 μM AlexaFluor 594. Somatic cell attached recordings were obtained from upper L2 neurons (50–200 μm depth from brain surface) visualized with the shadow patching technique[65]. Spikes were recorded either in current clamp or voltage clamp mode. Signals were filtered at 20 kHz (Multiclamp 700B, Axon Instruments) and digitized at 50 kHz using Wavesurfer (wavesurfer.janelia.org/). The frame trigger pulses of ScanImage were also recorded and used offline to synchronize individual frames to electrophysiological recordings. After establishment of a low-resistance seal (15–50 MΩ), randomized visual stimulation was delivered to increase the activity of the cells in the FOV. In a small subset of recordings, we microstimulated the recorded neuron in voltage clamp recording mode by applying DC current to increase its firing probability[66].

Visual stimulation

Visual stimuli were moving gratings generated using the Psychophysics Toolbox in MATLAB (Mathworks), presented using an LCD monitor (30 × 40 cm^2), placed 25 cm in front of the centre of the right eye of the mouse. Each stimulus trial consisted of a 2-s blank period (uniform grey display at

mean luminance) followed by a 2-s drifting sinusoidal grating (0.05 cycles per degree, temporal frequency of 1 Hz, eight randomized different directions). The stimuli were synchronized to individual image frames using frame-start pulses provided by ScanImage.

Post hoc anatomy of the mouse cortex

After the loose-seal recording sessions, mice were anaesthetized with a mixture of ketamine–xylazine (0.1 mg ketamine and 0.008 mg xylazine per gram body weight) and were transcardially perfused with 4% PFA in 1X Dulbecco's PBS (DPBS). The brains were extracted and post-fixed overnight in the perfusing solution. The brains were sectioned at 50-µm thickness, blocked with 2% BSA + 0.4 Triton X-100 (in PBS) for 1 h at room temperature, incubated with primary antibody (Rb-anti-GFP, 1:500; G10362, Invitrogen) for 2 days at 4 °C, and secondary antibody (AlexaFluor 594 conjugated goat anti-Rb, 1:500; A-11012, Invitrogen) overnight at 4 °C. The sections were mounted on microscope slides in Vectashield hard-set antifade mounting medium with DAPI (H-1500, Vector). Samples were imaged using a TissueFAXS 200 slide scanner (TissueGnostics) comprising an X-Light V2 spinning disk confocal imaging system (CrestOptics) built on an Axio Imager.Z2 microscope (Carl Zeiss Microscopy) equipped with a Plan-Apochromat ×20/0.8 M27 objective lens.

Analysis of two-photon imaging of the mouse cortex

The acquired data were analysed using MATLAB (population imaging) or Python (imaging during loose-seal recordings). In the MATLAB pipeline, for every recorded FOV, we selected ROIs covering all identifiable cell bodies using a semi-automated algorithm, and the fluorescence time course was measured by averaging all pixels within individual ROIs, after correction for neuropil contamination ($r = 0.7$), as previously described in detail[5]. We used one-way ANOVA tests ($P < 0.01$) for identifying cells with significant increase in their fluorescence signal during the stimulus presentation (responsive cells). We calculated $\Delta F/F_0 = (F - F_0)/F_0$, where F is the instantaneous fluorescence signal and F_0 is the average fluorescence in the interval 0.7 s before the start of the visual stimulus. For each responsive cell, we defined the preferred stimulus as the stimulus that evoked the

maximal $\Delta F/F_0$ amplitude (peak values during the 2 s of stimulus presentation). The half-decay time was calculated as follows: for each responsive cell, we averaged its $\Delta F/F_0$ response to the preferred stimulus over five trials. We also calculated the standard deviation of the averaged baseline signal over the 0.7 s before the start of the stimulus. Only cells for which the maximal $\Delta F/F_0$ amplitude was higher than four standard deviations above the baseline signal were included in the analysis. The time required for each trace to reach half of its peak value (baseline fluorescence subtracted) was calculated by linear interpolation. The fraction of cells detected as responsive was calculated as the number of significantly responsive cells over all the cells analysed. The cumulative distribution of peak $\Delta F/F_0$ responses included the maximal response amplitude from all analysed cells, calculated as described above for the preferred stimulus of each cell. The OSI was calculated as before[5,6] by fitting the fluorescence response from individual cells to the eight drifting grating stimuli with two Gaussians, centred at the preferred response angle (R_{pref}) and the opposite angle (R_{opp}). The OSI was calculated as:

$${\rm{OSI}}=\frac{{R}_{{\rm{pref}}}-{R}_{{\rm{orth}}}}{{R}_{{\rm{pref}}}+{R}_{{\rm{orth}}}}$$

where R_{orth} is the orthogonal angle to the preferred angle.

The movies recorded during loose-seal recordings were motion-corrected and segmented with the Python implementation of Suite2p (github.com/MouseLand/suite2p)[67]. The ROI corresponding to the loose-seal-recorded cell was then manually selected from the automatically segmented ROIs. For this dataset, we could calculate the neuropil contamination for most of the movies and got a distribution with a median of r_neu ~ 0.8 (Extended Data Fig. 14g–j), so we used this value uniformly for neuropil correction. Calcium events were defined by grouping APs with a 20-ms inclusion window. Then, we calculated $\Delta F/F_0 = (F - F_0)/F_0$, where F is the instantaneous fluorescence signal and F_0 was defined separately for all calcium events as the mean fluorescence value of the last 200 ms before the first AP in the group. Peak amplitudes were measured as the difference between the fluorescence intensity of the cell in the frame right before the

first AP in the group, and the fluorescence intensity of the cell in the frame after the 0–95% rise time of the calcium sensor relative to the last AP in the group (30 ms (jGCaMP7f), 30 ms (XCaMP-Gf), 5 ms (jGCaMP8f), 5 ms (jGCaMP8m) and 10 ms (jGCaMP8s)). The sensitivity index (*d′*) was calculated as:

$${d}^{{\prime} }=\frac{\bar{{A}}}{\sqrt{\frac{1}{2}{\sigma }^{2}\left(A\right)}}$$

where *A* is the amplitudes of isolated calcium transients induced by a single AP. We also calculated a global $\Delta F/F_0$ trace $(\Delta F/F_0)_{global}$, in which we used the 20th percentile of the fluorescence trace in a 60-s long running window as the $F_{0,global}$. In the analyses, we only included calcium events in which this $(\Delta F/F_0)_{global}$ value was less than 0.5 right before the AP, to include only events starting near baseline fluorescence values, to exclude non-linear summation and saturation. Traces in Fig. 4 were filtered with a Gaussian kernel ($\sigma = 5$ ms).

Mouse surgeries for the cerebellum

Young adult (postnatal day 42–98) male C57BL/6J (Jackson Labs) mice were anaesthetized using isoflurane (2.5% for induction and 1.5% during surgery). A circular craniotomy (diameter of 3 mm) above medial crus I (2 mm left and 1 mm posterior to the midline junction of the interparietal and occipital bones). Viral suspension (200 nl) was injected in two locations near the centre point at a depth of 300–400 µm. Constructs injected included: AAV2/1-*CAG*-FLEx-jGCaMP8 constructs (pGP-AAV-CAG-FLEx-jGCaMP8f-WPRE, Addgene plasmid #162382; pGP-AAV-CAG-FLEx-jGCaMP8m-WPRE, Addgene plasmid #162381; pGP-AAV-CAG-FLEx-jGCaMP8s-WPRE, Addgene plasmid #162380; pGP-AAV-CAG-FLEx-jGCaMP7f-WPRE, Addgene plasmid #104496; and pGP-AAV-CAG-FLEx-jGCaMP6f-WPRE, Addgene plasmid #100835; all viruses were diluted to 4×10^{12} GC per millilitre titre).

Purkinje cell-specific expression was induced by co-injection of virus expressing Cre under control of a promoter fragment from the Purkinje cell protein 2 (*Pcp2*; also known as *L7*) gene (AAV2/1-*sL7*-Cre, 5.3×10^{10} GC

per millilitre titre)[68]. A 3-mm diameter circular coverslip glued to a donut-shaped 3.5-mm diameter coverslip (no. 1 thickness, Warner Instruments) was cemented to the craniotomy using dental cement (C&B Metabond, Parkell). A custom titanium head post was cemented to the skull.

Two-photon imaging of the cerebellum

Head-restrained mice were allowed to freely locomote on a wheel. Imaging was performed with a custom-built two-photon microscope with a resonant scanner. The light source was a Mai Tai Sapphire laser (Spectra Physics) running at 920 nm. The objective was a ×16 CFI LWD Plan fluorite objective water immersion lens with 0.8 NA (Nikon). The detection path consisted of a bandpass filter (525/50 nm) and a 565-nm dichroic mirror directed towards a photomultiplier tube (Hamamatsu). Images were acquired using ScanImage (vidriotechnologies.com)[64]. Functional images (512 × 32 pixels, 215 × 27 μm^2) of Purkinje cell dendrites (50–250 μm below the pia mater) were collected at 283 Hz. Laser power was up to 50 mW at the front aperture of the objective.

Analysis of two-photon imaging in the cerebellum

Purkinje cell dendrite movies were captured during free locomotion without applied stimulation. Movies were motion-corrected and converted to $\Delta F/F_0$ traces using the Python implementation of CaImAn[69]. Individual events within the traces were identified by finding adjacent local maxima in $\Delta F/F_0$ variance that had one local maximum in the $\Delta F/F_0$ trace between them. Statistics for individual events were calculated by fitting the equation:

$$F(t)=({F}_{\max }{(1-{e}^{-(t-{t}_{0})/{\tau }_{{\rm{rise}}}})}^{a}+{F}_{{\rm{start}}}/({F}_{1}+{F}_{2}))({F}_{1}{e}^{-(t-{t}_{0})/{\tau }_{1}}+{F}_{2}{e}^{-(t-{t}_{0})/{\tau }_{2}})$$

where t_0 is the start of the peak, τ_{rise} is the rise time constant shaped by α; τ_1 and τ_2 are the decay time constants, F_{max} is the maximum amplitude of the trace above the starting point F_{start}, and F_1 and F_2 are component amplitudes.

Spike fluorescence models

We used a phenomenological model that converts spike times to a synthetic fluorescence time series[33]. This S2F model consists of two steps. First, spikes at times $\{t_k\}$ are converted to a latent variable, $c(t)$, by convolution with two double-exponential kernels:

$$c(t)={\sum }_{t > {t}_{k}}\left[r\bullet \exp (\,-\,\frac{t-{t}_{k}}{{\tau }_{d1}})+\exp (\,-\,\frac{t-{t}_{k}}{{\tau }_{d2}})\right]\bullet \left[1-\exp (\,-\,\frac{t-{t}_{k}}{{\tau }_{r}})\right]+{n}_{i}(t)$$

$\tau_{\rm r}$, and $\tau_{\rm d1}$ and $\tau_{\rm d2}$ are the rise time and decay times, respectively. In our model, we required $\tau_{\rm d1} < \tau_{\rm d2}$ (that is, fast and slow components), with r representing the ratio of the weight for the fast component to that of the slow one. ${n}_{i}\left(t\right) \sim N\left(0,{\sigma }_{i}^{2}\right)$ is the Gaussian-distributed noise. $c(t)$ was truncated at zero if noise drove it to negative values. We tested the performance of models with various choices of rise and decay times: (1) one rise and one decay time, (2) one rise and two decay times, and (3) two rise and two decay times. Using cross-validation, we found that the models for one rise and two decay times fit pyramidal cells (as described above), whereas interneurons were fit well by one rise and a single decay time (as described above with $r = 0$). Subsequently, $c(t)$ was converted to a synthetic fluorescence signal through a sigmoidal function:

$$\Delta F\,/\,{F}_{{\rm{Synth}}}(t)=\frac{{F}_{m}}{1+\exp [-k(c(t)-{c}_{1/2})]}+{n}_{e}(t)$$

where k is a non-linearity sharpness parameter, $c_{1/2}$ is a half-activation parameter and F_m is the maximum possible fluorescence change. ${n}_{e}\left(t\right) \sim N(0,{\sigma }_{e}^{2})$ is the Gaussian-distributed external noise. For comparison, we also generated a S2F linear model with $\,\Delta F/\,{F}_{{\rm{Synth}}}(t)=$ ${F}_{\max }c\left(t\right)+{F}_{0}$, where $F_{\max}$ is a scaling parameter (we kept the naming as max to clarify the relationship to other models); F_0 is the baseline.

We applied the linear[33] (Fig. 5d–h) and non-linear[70] (Extended Data Fig. 19a–e) inference models using their default parameters. The linear model provided an estimate of the number of spikes per imaging frame (Fig. 5d, middle panel), in which we considered that a spiking event occurred if the estimated spike count was above 0.6. The non-linear model provided an estimate of spike times (inferred spikes). Both models provide reconstructions of fluorescent dynamics $\Delta F/F_{\rm Synth}$, based on inferred spikes and estimated model parameters from spikes to fluorescence. We computed spiking dynamics based on spike counts in a frame convolved with a Gaussian filter with standard deviation of 100 ms for both ground-truth (Fig. 5d, bottom, black line, $r(t)$) and inferred spikes (Fig. 5d, bottom, red line, $\hat{r}(t)$).

We measured the inference performance using (1) variance explained of fluorescence dynamics as $1-\frac{{ < {(\Delta F/F(t)-\Delta F/{F}_{{\rm{Synth}}}(t))}^{2} > }_{t}}{{ < {(\Delta F/F(t)-{ < \Delta F/F(t) > }_{t})}^{2} > }_{t}}$ (Fig. 5e); (2) variance explained of spiking dynamics as $1-\frac{{ < {({\rm{r}}(t)-\hat{{\rm{r}}}(t))}^{2} > }_{t}}{{ < {({\rm{r}}(t)-{ < {\rm{r}}(t) > }_{t})}^{2} > }_{t}}$ (Fig. 5f); (3) spike detectability ${F}_{{score}}=\frac{2\times \#\{{true\; in\; ferred\; spikes}\}}{\#\{{true\; spikes}\}+\#\{{in\; ferred\; spikes}\}}$, where an inferred spike falling into a 40-ms time window of a true spike (that is, ${|t}_{{\rm{true}}}-{t}_{{\rm{inferred}}}| < 40\,{\rm{ms}}$) was considered as a true inferred spike (true positive) (Fig. 5g); and (4) for true inferred spikes, we measured spike timing error as $\sqrt{ < {\left({t}_{{\rm{true}}}-{t}_{{\rm{inferred}}}\right)}^{2}\, > }$ (Fig. 5h).

Variance explained

Variance explained measures the goodness-of-fit of an S2F model, as $1-\frac{{(\Delta F/{F}_{{\rm{raw}}}(t)-\Delta F/{F}_{{\rm{Synth}}}(t))}^{2}}{{(\Delta F/{F}_{{\rm{raw}}}(t)-{ < \Delta F/{F}_{{\rm{raw}}}(t) > }_{t})}^{2}}$. Here we used only the time t with a spike rate of more than 0 Hz after spikes in the calculation, in which the instantaneous spike rate at time t is estimated by a boxcar-rolling average over a 600-ms time window.

Experimental design

Sample size

For the cultured neuron assay, we used *n* consistent with our power analysis from an earlier study[24]; prioritized variants received many more replicates (11, 24 and 64 for the three jGCaMP8 indicators). We found the purified protein experiments to be extremely reproducible, with *n* of 3–5 routinely providing very small error bars. Many individuals were used for each in vivo experiment, with multiple trials per FOV per individual; values of *n* ranged into the hundreds for these experiments. All multiple-comparison experiments were verified to provide sufficient power with the Kruskal–Wallis multiple-comparison test before proceeding to pairwise comparison tests.

Randomization

There were no experimental groups, thus no need for randomization.

Blinding

There were no experimental groups, thus no need for blinding.

Statistics

Exact statistical tests used for each comparison, as well as *n*, are listed in the main text and figure legends. Box-and-whisker plots throughout the paper indicate the median and 25–75th percentile range; whiskers indicate the shorter of 1.5 times the interquartile range or the extreme data point[71].

For Fig. 3e, full statistics are: jGCaMP7f (min, Q1, Q2, Q3 and max) = 0.013, 0.041, 0.11, 0.22 and 1.16, $n = 320$ cells from 3 mice; XCaMP-Gf (min, Q1, Q2, Q3 and max) = 0.016, 0.048, 0.091, 0.16 and 0.38, $n = 124$ cells from 3 mice; jGCaMP8f (min, Q1, Q2, Q3 and max) = 0.010, 0.033, 0.084, 0.15 and 0.44, $n = 317$ cells from 5 mice; jGCaMP8m (min, Q1, Q2, Q3 and max) = 0.011, 0.032, 0.084, 0.16 and 0.56, $n = 365$ cells from 3 mice; jGCaMP8s (min, Q1, Q2, Q3 and max) = 0.013, 0.059, 0.14, 0.24 and

0.67, $n = 655$ cells from 6 mice. Kruskal–Wallis multiple-comparison test, $P = 3.1 \times 10^{-15}$. jGCaMP7f versus XCaMP-Gf: $P = 1.0$; jGCaMP7f versus jGCaMP8f: $P = 0.013$; jGCaMP7f versus jGCaMP8m: $P = 0.029$; jGCaMP7f versus jGCaMP8m: $P = 0.010$; jGCaMP7f versus jGCaMP8s: $P = 0.010$; XCaMP-Gf versus jGCaMP8f: $P = 1.0$; XCaMP-Gf versus jGCaMP8m: $P = 1.0$; XCaMP-Gf versus jGCaMP8s: $P = 0.0027$; jGCaMP8f versus jGCaMP8m: $P = 1.0$; jGCaMP8f versus jGCaMP8s: $P = 2.4 \times 10^{-11}$; and jGCaMP8m versus jGCaMP8s: $P = 4.49 \times 10^{-11}$. For Dunn's comparison test shown: *$P < 0.05$, ***$P < 0.001$ and not significant. For Fig. 3f, data passed Shapiro–Wilk normality test ($\alpha = 0.05$ level). For Fig. 3f, full statistics are: jGCaMP8f (min, Q1, Q2, Q3 and max) = 0.12, 0.22,0.375, 0.52 and 0.72, $n = 19$ FOVs from 5 mice; jGCaMP8m (min, Q1, Q2, Q3 and max) = 0.32, 0.55, 0.66, 0.79 and 0.85, $n = 14$ FOVs from 3 mice; jGCaMP8s (min, Q1, Q2, Q3 and max) = 0.33, 0.71, 0.79, 0.89 and 1.00, $n = 26$ FOVs from 6 mice; and jGCaMP7f (min, Q1, Q2, Q3 and max) = 0.21, 0.32, 0.45, 0.59 and 0.73, $n = 12$ FOVs from 3 mice. For Tukey's multiple comparison test: jGCaMP7f versus jGCaMP8f: $P = 0.83$; jGCaMP7f versus jGCaMP8m: $P = 0.0184$; jGCaMP7f versus jGCaMP8s: $P = 4.52 \times 10^{-5}$; jGCaMP8f versus jGCaMP8m: $P = 0.00098$; and jGCaMP8m versus jGCaMP8s: $P = 0.23$. For one-way ANOVA test: *$P < 0.05$, ***$P < 0.001$ and not significant. For Fig. 3g, full statistics are: jGCaMP7f versus jGCaMP8f: $P = 8.7 \times 10^{-103}$; jGCaMP7f versus jGCaMP8m: $P = 1.0$; jGCaMP7f versus jGCaMP8s: $P = 5.5 \times 10^{-85}$; jGCaMP8f versus jGCaMP8m: $P = 9.3 \times 10^{-96}$; and jGCaMP8m versus jGCaMP8s: $P = 1.1 \times 10^{-72}$ (Kruskal–Wallis test with Dunn's multiple comparison test was used to compare the magnitude of response across groups).

For Extended Data Fig. 8e, at a duration of 4 ms, Kruskal–Wallis test found $P = 2.3 \times 10^{-3}$ and pairwise Dunn's multiple comparison test to jGCaMP7f as follows: jGCaMP8f $P = 0.03$, jGCaMP8m $P = 2.0 \times 10^{-4}$ and jGCaMP8s $P = 0.24$. At a duration of 8 ms, Kruskal–Wallis test found $P = 3.5 \times 10^{-5}$ and pairwise Dunn's multiple comparison test to jGCaMP7f as follows: jGCaMP8f $P = 3.5 \times 10^{-3}$, jGCaMP8m $P = 2.8 \times 10^{-5}$ and jGCaMP8s $P = 0.73$. At a duration of 25 ms, Kruskal–Wallis test found $P = 3.4 \times 10^{-5}$ and pairwise Dunn's multiple comparison test to jGCaMP7f as follows:

jGCaMP8f $P = 0.074$, jGCaMP8m $P = 1.6 \times 10^{-3}$ and jGCaMP8s $P = 0.11$. Numbers tested are the same as Extended Data Fig. 8c.

Reagent availability

DNA constructs and AAV particles of jGCaMP8s, jGCaMP8m and jGCaMP8f (pCMV, pAAV-synapsin-1, pAAV-synapsin-1-FLEX and pAAV-CAG-FLEX) have been deposited at Addgene (#162371–162382). Sequences have been deposited in GenBank (#OK646318–OK646320). The crystal structure of jGCaMP8.410.80 has been deposited in the PDB (PDB ID: 7ST4). *Drosophila* stocks were deposited at the Bloomington Drosophila Stock Center (http://flystocks.bio.indiana.edu) - #92587–92595; *Drosophila* UAS and lexAOp plasmids have been deposited at Addgene (#162383–162388). Email GENIEreagents@janelia.hhmi.org for additional requests.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Most datasets generated for characterizing the new sensors are included in the Article (and its Supplementary Information files). In vivo mouse cell-attached datasets are available on the DANDI Archive (https://dandiarchive.org/dandiset/000168). Additional datasets are available from the corresponding authors on reasonable request.

Code availability

The code for analysing the neuron culture screening results is available at https://github.com/ilyakolb/jGCaMP8-neuron-culture-screen. The custom code for the S2F model is available at https://github.com/zqwei/Spike2Fluorescence_jGCaMP8. The example

python code for usage of the in vivo mouse cell-attached dataset is available at https://github.com/rozmar/jGCaMP8_ground_truth_dataset.

References

1. Miyawaki, A. et al. Fluorescent indicators for Ca^{2+} based on green fluorescent proteins and calmodulin. *Nature* **388**, 882–887 (1997).

2. Nakai, J., Ohkura, M. & Imoto, K. A high signal-to-noise Ca^{2+} probe composed of a single green fluorescent protein. *Nat. Biotechnol.* **19**, 137–141 (2001).

3. Tian, L. et al. Imaging neural activity in worms, flies and mice with improved GCaMP calcium indicators. *Nat. Methods* **6**, 875–881 (2009).

4. Akerboom, J. et al. Optimization of a GCaMP calcium indicator for neural activity imaging. *J. Neurosci. Off. J. Soc. Neurosci.* **32**, 13819–13840 (2012).

5. Chen, T.-W. et al. Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature* **499**, 295–300 (2013).

6. Dana, H. et al. High-performance calcium sensors for imaging activity in neuronal populations and microcompartments. *Nat. Methods* **16**, 649–657 (2019).

7. Inoue, M. et al. Rational design of a high-affinity, fast, red calcium indicator R-CaMP2. *Nat. Methods* **12**, 64–70 (2015).

8. Inoue, M. et al. Rational engineering of XCaMPs, a multicolor GECI suite for in vivo imaging of complex brain circuit dynamics. *Cell* **177**, 1346–1360.e24 (2019).

9. Grienberger, C. & Konnerth, A. Imaging calcium in neurons. *Neuron* **73**, 862–885 (2012).

10. Luo, L., Callaway, E. M. & Svoboda, K. Genetic dissection of neural circuits: a decade of progress. *Neuron* **98**, 256–281 (2018).

11. Sofroniew, N. J., Flickinger, D., King, J. & Svoboda, K. A large field of view two-photon mesoscope with subcellular resolution for in vivo imaging. *eLife* **5**, e14472 (2016).

12. Ahrens, M. B., Orger, M. B., Robson, D. N., Li, J. M. & Keller, P. J. Whole-brain functional imaging at cellular resolution using light-sheet microscopy. *Nat. Methods* **10**, 413–420 (2013).

13. Skocek, O. et al. High-speed volumetric imaging of neuronal activity in freely moving rodents. *Nat. Methods* **15**, 429–432 (2018).

14. Petreanu, L. et al. Activity in motor–sensory projections reveals distributed coding in somatosensation. *Nature* **489**, 299–303 (2012).

15. Maravall, M., Mainen, Z. F., Sabatini, B. L. & Svoboda, K. Estimating intracellular calcium concentrations and buffering without wavelength ratioing. *Biophys. J.* **78**, 2655–2667 (2000).

16. Sabatini, B. L. & Regehr, W. G. Timing of neurotransmission at fast synapses in the mammalian brain. *Nature* **384**, 170–172 (1996).

17. Svoboda, K., Denk, W., Kleinfeld, D. & Tank, D. W. In vivo dendritic calcium dynamics in neocortical pyramidal neurons. *Nature* **385**, 161–165 (1997).

18. Sabatini, B. L., Oertner, T. G. & Svoboda, K. The life cycle of Ca^{2+} ions in dendritic spines. *Neuron* **33**, 439–452 (2002).

19. Tay, L. H., Griesbeck, O. & Yue, D. T. Live-cell transforms between Ca^{2+} transients and FRET responses for a troponin-C-based Ca^{2+} sensor. *Biophys. J.* **93**, 4031–4040 (2007).

20. Sun, X. R. et al. Fast GCaMPs for improved tracking of neuronal activity. *Nat. Commun.* **4**, 2170 (2013).

21. Peron, S. P., Freeman, J., Iyer, V., Guo, C. & Svoboda, K. A cellular resolution map of barrel cortex activity during tactile behavior. *Neuron* **86**, 783–799 (2015).

22. Stringer, C., Pachitariu, M., Steinmetz, N., Carandini, M. & Harris, K. D. High-dimensional geometry of population responses in visual cortex. *Nature* **571**, 361–365 (2019).

23. Lukas, T. J., Burgess, W. H., Prendergast, F. G., Lau, W. & Watterson, D. M. Calmodulin binding domains: characterization of a phosphorylation and calmodulin binding site from myosin light chain kinase. *Biochemistry* **25**, 1458–1464 (1986).

24. Wardill, T. J. et al. A neuron-based screening platform for optimizing genetically-encoded calcium indicators. *PLoS ONE* **8**, e77728 (2013).

25. Pologruto, T. A., Yasuda, R. & Svoboda, K. Monitoring neural activity and [Ca^{2+}] with genetically encoded Ca^{2+} indicators. *J. Neurosci.* **24**, 9572–9579 (2004).

26. McNicol, D. *A Primer of Signal Detection Theory* (Psychology Press, 2004).

27. Strother, J. A., Nern, A. & Reiser, M. B. Direct observation of ON and OFF pathways in the *Drosophila* visual system. *Curr. Biol.* **24**, 976–983 (2014).

28. Juusola, M., Dau, A., Zheng, L. & Rien, D. Electrophysiological method for recording intracellular voltage responses of *Drosophila* photoreceptors and interneurons to light stimuli in vivo. *J. Vis. Exp.* https://doi.org/10.3791/54142 (2016).

29. Yang, H. H. et al. Subcellular imaging of voltage and calcium signals reveals neural processing in vivo. *Cell* **166**, 245–257 (2016).

30. Kerlin, A. M., Andermann, M. L., Berezovskii, V. K. & Reid, R. C. Broadly tuned response properties of diverse inhibitory neuron subtypes in mouse visual cortex. *Neuron* **67**, 858–871 (2010).

31. Yu, J., Hu, H., Agmon, A. & Svoboda, K. Recruitment of GABAergic interneurons in the barrel cortex during active tactile behavior. *Neuron* **104**, 412–427.e4 (2019).

32. Miyakawa, H., Lev-Ram, V., Lasser-Ross, N. & Ross, W. N. Calcium transients evoked by climbing fiber and parallel fiber synaptic inputs in guinea pig cerebellar Purkinje neurons. *J. Neurophysiol.* **68**, 1178–1189 (1992).

33. Wei, Z. et al. A comparison of neuronal population dynamics measured with calcium imaging and electrophysiology. *PLoS Comput. Biol.* **16**, e1008198 (2020).

34. Pnevmatikakis, E. A. et al. Simultaneous denoising, deconvolution, and demixing of calcium imaging data. *Neuron* **89**, 285–299 (2016).

35. Vogelstein, J. T. et al. Fast nonnegative deconvolution for spike train inference from population calcium imaging. *J. Neurophysiol.* **104**, 3691–3704 (2010).

36. Steinmetz, N. A. et al. Aberrant cortical activity in multiple GCaMP6-expressing transgenic mouse lines. *eNeuro* **4**, ENEURO.0207-17.2017 (2017).

37. Picardo, M. A. et al. Population-level representation of a temporal sequence underlying song production in the zebra finch. *Neuron* **90**, 866–876 (2016).

38. Kerr, J. N. D., Greenberg, D. & Helmchen, F. Imaging input and output of neocortical networks in vivo. *Proc. Natl Acad. Sci. USA* **102**, 14063–14068 (2005).

39. Peron, S., Chen, T.-W. & Svoboda, K. Comprehensive imaging of cortical networks. *Curr. Opin. Neurobiol.* **32**, 115–123 (2015).

40. Chen, Y. et al. Soma-targeted imaging of neural circuits by ribosome tethering. *Neuron* **107**, 454–469.e6 (2020).

41. Chamberland, S. et al. Fast two-photon imaging of subcellular voltage dynamics in neuronal tissue with genetically encoded indicators. *eLife* **6**, e25690 (2017).

42. Ramirez, J. E. & Stell, B. M. Calcium imaging reveals coordinated simple spike pauses in populations of cerebellar Purkinje cells. *Cell Rep.* **17**, 3125–3132 (2016).

43. Stanford, T. R., Shankar, S., Massoglia, D. P., Costello, M. G. & Salinas, E. Perceptual decision making in less than 30 milliseconds. *Nat. Neurosci.* **13**, 379–385 (2010).

44. Baird, G. S., Zacharias, D. A. & Tsien, R. Y. Circular permutation and receptor insertion within green fluorescent proteins. *Proc. Natl Acad. Sci. USA* **96**, 11241–11246 (1999).

45. Barykina, N. V. et al. Green fluorescent genetically encoded calcium indicator based on calmodulin/M13-peptide from fungi. *PLoS ONE* **12**, e0183757 (2017).

46. Barykina, N. V. et al. FGCaMP7, an improved version of fungi-based ratiometric calcium indicator for in vivo visualization of neuronal activity. *Int. J. Mol. Sci.* **21**, 3012 (2020).

47. Akerboom, J. et al. Genetically encoded calcium indicators for multi-color neural activity imaging and combination with optogenetics. *Front. Mol. Neurosci.* **6**, 2 (2013).

48. Tallini, Y. N. et al. Imaging cellular signals in the heart in vivo: cardiac expression of the high-signal Ca^{2+} indicator GCaMP2. *Proc. Natl Acad. Sci. USA* **103**, 4753–4758 (2006).

49. Abdelfattah, A. S. et al. Bright and photostable chemigenetic indicators for extended in vivo voltage imaging. *Science* **365**, 699–704 (2019).

50. Xu, C. & Webb, W. W. Measurement of two-photon excitation cross sections of molecular fluorophores with data from 690 to 1050 nm. *J.*

Opt. Soc. Am. B **13**, 481–491 (1996).

51. Makarov, N. S., Drobizhev, M. & Rebane, A. Two-photon absorption standards in the 550–1600 nm excitation wavelength range. *Opt. Express* **16**, 4029–4047 (2008).

52. Mütze, J. et al. Excitation spectra and brightness optimization of two-photon excited probes. *Biophys. J.* **102**, 934–944 (2012).

53. Bengtson, C. P. & Bading, H. in *Synaptic Plasticity: Dynamics, Development and Disease* (eds. Kreutz, M. R. & Sala, C.) 377–405 (Springer, 2012).

54. Berg, S. et al. ilastik: Interactive machine learning for (bio)image analysis. *Nat. Methods* **16**, 1226–1232 (2019).

55. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).

56. Akerboom, J. et al. Crystal structures of the GCaMP calcium sensor reveal the mechanism of fluorescence signal change and aid rational design. *J. Biol. Chem.* **284**, 6455–6464 (2009).

57. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240–255 (1997).

58. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).

59. Dionne, H., Hibbard, K. L., Cavallaro, A., Kao, J.-C. & Rubin, G. M. Genetic reagents for making split-GAL4 lines in *Drosophila*. *Genetics* **209**, 31–35 (2018).

60. Pfeiffer, B. D., Truman, J. W. & Rubin, G. M. Using translational enhancers to increase transgene expression in *Drosophila*. *Proc. Natl Acad. Sci. USA* **109**, 6626–6631 (2012).

61. Aso, Y. et al. The neuronal architecture of the mushroom body provides a logic for associative learning. *eLife* **3**, e04577 (2014).

62. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).

63. Goldey, G. J. et al. Removable cranial windows for long-term imaging in awake mice. *Nat. Protoc.* **9**, 2515–2538 (2014).

64. Pologruto, T. A., Sabatini, B. L. & Svoboda, K. ScanImage: flexible software for operating laser scanning microscopes. *Biomed. Eng. Online* **2**, 13 (2003).

65. Kitamura, K., Judkewitz, B., Kano, M., Denk, W. & Häusser, M. Targeted patch-clamp recordings and single-cell electroporation of unlabeled neurons in vivo. *Nat. Methods* **5**, 61–67 (2008).

66. Perkins, K. L. Cell-attached voltage-clamp and current-clamp recording and stimulation techniques in brain slices. *J. Neurosci. Methods* **154**, 1–18 (2006).

67. Pachitariu, M. et al. Suite2p: beyond 10,000 neurons with standard two-photon microscopy. Preprint at https://doi.org/10.1101/061507 (2017).

68. Oberdick, J., Smeyne, R. J., Mann, J. R., Zackson, S. & Morgan, J. I. A promoter that drives transgene expression in cerebellar Purkinje and retinal bipolar neurons. *Science* **248**, 223–226 (1990).

69. Giovannucci, A. et al. CaImAn an open source tool for scalable calcium imaging data analysis. *eLife* **8**, e38173 (2019).

70. Deneux, T. et al. Accurate spike estimation from noisy calcium signals for ultrafast three-dimensional imaging of large neuronal populations in vivo. *Nat. Commun.* **7**, 12190 (2016).

71. Krzywinski, M. & Altman, N. Visualizing samples with box plots. *Nat. Methods* **11**, 119–120 (2014).

Acknowledgements

This work was done as a collaboration with the Janelia Genetically Encoded Indicator and Effector Project Team (GENIE; https://www.janelia.org/project-team/genie) at the Howard Hughes Medical Institute, Janelia Research Campus. J.Z., D.R., A.T., G.T., J.P.H. and I.K. are members of GENIE. G.C.T., W.L.K., E.R.S., K.S., J.P.H., I.K. and L.L.L. are on the GENIE Steering Committee. We thank additional GENIE Steering Committee members V. Jayaraman, N. Spruston and A. Tebo for useful discussions. The work was supported by the Howard Hughes Medical Institute, NIH U19 NS104648 and NIH R01 NS045193 to S.S.-H.W., and NIH F32 MH120887 to G.J.B. We thank the Berkeley Center for Structural Biology for use of beamline 8.2.1; and are grateful to the Viral Tools, Cell and Tissue Culture, Molecular Biology, Media Prep, Vivarium/Aquarium, Anatomy and Histology, and Fly Facility Shared Resources at Janelia for technical assistance with numerous parts of the project.

Author information

Author notes

1. These authors contributed equally: Yan Zhang, Márton Rózsa

Authors and Affiliations

1. Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA, USA

   Yan Zhang, Márton Rózsa, Yajie Liang, Daniel Bushey, Ziqiang Wei, Jihong Zheng, Daniel Reep, Arthur Tsang, Getahun Tsegaye, Sujatha Narayan, Christopher J. Obara, Jing-Xuan Lim, Ronak Patel, Rongwei Zhang, Misha B. Ahrens, Glenn C. Turner, Wyatt L. Korff, Eric R. Schreiter, Karel Svoboda, Jeremy P. Hasseman, Ilya Kolb & Loren L. Looger

2. Allen Institute for Neural Dynamics, Seattle, WA, USA

   Márton Rózsa, Sujatha Narayan & Karel Svoboda

3. Genetically Encoded Neural Indicator and Effector (GENIE) Project, Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA, USA

   Yajie Liang, Jihong Zheng, Daniel Reep, Arthur Tsang, Getahun Tsegaye, Glenn C. Turner, Wyatt L. Korff, Eric R. Schreiter, Karel Svoboda, Jeremy P. Hasseman, Ilya Kolb & Loren L. Looger

4. Department of Diagnostic Radiology and Nuclear Medicine, University of Maryland School of Medicine, Baltimore, MD, USA

   Yajie Liang

5. Neuroscience Institute, Princeton University, Princeton, NJ, USA

   Gerard Joey Broussard & Samuel S.-H. Wang

6. Department of Neurosciences, University of California, San Diego, La Jolla, CA, USA

   Loren L. Looger

Contributions

Y.Z., E.R.S. and L.L.L. designed the project. Y.Z., K.S., W.L.K., I.K., J.P.H. and L.L.L. led the project. Y.Z. and E.R.S. optimized peptides for grafting. Y.Z., G.T. and J.P.H. performed mutagenesis. D.B., J.Z. and G.C.T. contributed flies. G.J.B. and S.S.-H.W. contributed cerebellum. Y.Z., R.P. and I.K. performed photophysics. Y.Z. performed crystallography and purified protein experiments. M.R., Y.L. and K.S. contributed cortex. M.R. and K.S. performed simultaneous imaging and electrophysiology. Z.W., M.R. and K.S. performed S2F and F2S. I.K., J.P.H. and K.S. contributed the cultured neuron screen design. Y.L., D.R., A.T., C.J.O., R.Z., J.P.H. and I.K. performed cultured neuron experiments. D.R. and A.T. performed histology.

Y.Z., M.R., Y.L., K.S., I.K., D.B., J.Z., Z.W., J.-X.L., R.Z., M.B.A. and S.N. performed the analysis methodologies. Y.Z. and J.P.H. coordinated shipments and deposits. Y.Z., M.R., Y.L., K.S., D.B., I.K. and L.L.L. wrote the paper, with contributions from all the authors.

Corresponding authors

Correspondence to Glenn C. Turner, Samuel S.-H. Wang, Karel Svoboda, Jeremy P. Hasseman or Loren L. Looger.

Ethics declarations

Competing interests

Y.Z., E.R.S., J.P.H., I.K., K.S. and L.L.L. are inventors of US Patent Application 63082222, 'Genetically Encoded Calcium Indicators and Methods of Use', which covers the jGCaMP8 sensors and is assigned to HHMI. The remaining authors declare no competing interests.

Peer review

Peer review information

Nature thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Crystal structure of jGCaMP8.410.80.

ENOSP (yellow), linker 1 (ENOSP-cpGFP, grey), linker 2 (cpGFP-CaM, grey), cpGFP (green), CaM (blue), Ca^{2+} ions (orange). a. Overlay of the structures of jGCaMP8.410.80 and GCaMP5G (light grey). Left: side view. Right: top view. b. A closeup of the chromophore region in structures of jGCaMP8.410.80 and GCaMP5G. Ile32 (dark gray) in Linker 1 of jGCaMP8.410.80 facilitates closer interaction of Tyr352 (blue) with the GFP chromophore. The corresponding residues in GCaMP5G, Glu60 and Tyr380, are depicted in light gray. c. Individual residue mutations screened in this study, shown on the structure of jGCaMP8.410.80. Sixteen initial interface positions are in orange. Ten subsequently mutated CaM positions are in magenta. Mutations based on the FGCaMP sensor are in cyan.

Extended Data Fig. 2 Results of cultured neuron 1-AP field stimulation screen (*n* = 813 constructs, 647 with detectable 1-AP responses; Methods).

All results are normalized to in-plate GCaMP6s controls (blue line) and listed in ranked order (increasing for peak $\Delta F/F_0$, decreasing for all others) from each screening round. Other relevant control constructs (n = 9) were screened side-by-side (right panels). Sensor engineering took place over seven rounds: Round 0 (r0): Graft peptides (n = 29 constructs). Round 1 (r1): Screen linkers (n = 64 constructs). Round 2 (r2): Site-saturation mutagenesis of 16 interface positions: 7 in ENOSP, 4 on cpGFP, and 5 on CaM (n = 304 constructs). Round 3 (r3): Combination of beneficial mutations to date (n = 69 constructs). Round 4 (r4): Site-saturation mutagenesis of 10 additional CaM positions surrounding ENOSP and of 3 residues on linker1 (n = 272 constructs). Graft mutations from FGCaMP. Round 5 (r5) and 6 (r6): Two additional rounds of combination of beneficial mutations (n = 25, 51 constructs respectively).

Extended Data Fig. 3 Response characteristics of jGCaMP8 indicators to 3, 10, and 160 field stimulation pulses (45 V, 83 Hz).

Half-decay at 160 pulses is not reported because cell fluorescence typically does not decay to baseline during our imaging time (6 s after stimulus

onset). *n* values same as in Fig. 1d,e. Box-whisker plots indicate the median and 25th–75th percentile range; whiskers indicate the shorter of 1.5 times the inter-quartile range or the extreme data point.

Extended Data Fig. 4 Baseline brightness and photobleaching of sensors.

a. Baseline brightness. The jGCaMP8 series exhibited similar baseline fluorescence in the cultured neuron assay compared to jGCaMP7f, but XCaMP sensors were significantly dimmer ($H(6) = 71.77$, $P < 0.0001$, Kruskal-Wallis test; Dunn's multiple comparisons test with jGCaMP7f as control). n.s.: not significant ($P > 0.99$). *$P = 0.012$; **$P = 0.0012$; ****$P < 0.0001$. Each point represents median neuronal brightness from a single well. jGCaMP8f: n = 40, jGCaMP8m: n = 8, jGCaMP8s: n = 18, jGCaMP7f: n = 20, XCaMP-Gf: n = 29, XCaMP-G: n = 31, XCaMP-Gf0: n = 16; overall statistics: n = 2 independent transfections, 5 96-well plates. Box-whisker plots indicate the median and 25th–75th percentile range; whiskers indicate the shorter of 1.5 times the inter-quartile range or the extreme data point. b. Photobleaching of jGCaMP8, jGCaMP7, and XCaMP variants in neuron cell culture. Grey lines: individual cells, black lines: mean. Each cell's fluorescence trace was normalized to the initial value. *N* values indicate number of cells ($n = 1$ well per variant, $n = 1$ transfection day). After continuous illumination for 10 min, neurons transfected with jGCaMP8 variants lost on average 13-28% of their initial fluorescence. jGCaMP8m exhibited biphasic bleaching: a rapid phase consisting of ~15% fluorescence loss within 10 s followed by a slower phase (10% within 10 min). Of the other variants, jGCaMP7c also exhibited this property. We noticed considerable variability in the photobleaching rates within individual neurons, possibly stemming from expression level and differences in baseline brightness in each neuron as a function of intracellular resting [Ca^{2+}].

Extended Data Fig. 5 Photophysical characterization of jGCaMP8 sensors.

a. One-photon absorbance spectra of jGCaMP sensors acquired in 10 mM MOPS, pH 7.2. b. One-photon excitation and emission spectra of jGCaMP8 sensors. Emission spectra were calculated with 460 nm excitation light (bandwidth 5 nm); excitation spectra were calculated with 540 nm emission light (bandwidth 5 nm). Averaged data from $n = 2$ independent measurements per sensor. c. Two-photon action cross-sections of jGCaMP8 sensors. Averaged data from $n = 2$ independent measurements per sensor. d. Molecular brightness. Averaged data from $n = 2$ independent measurements per sensor.

Extended Data Fig. 6 Linearity of $\Delta F/F_0$ of jGCaMP8, jGCaMP7, and XCaMP variants in cultured neurons.

Each gray dot represents a single well. $\Delta F/F_0$ values in the 1–10 and 1–40 pulse range were fit to a linear model (orange and green, respectively). The slopes (*m*) and R^2 values are reported for each fit. jGCaMP8f, 29 wells, 594 neurons; jGCaMP8m, 16 wells, 408 neurons; jGCaMP8s, 12 wells, 121 neurons; GCaMP6s, 14 wells, 187 neurons; jGCaMP7s, 14 wells, 177 neurons; jGCaMP7c, 13 wells, 117 neurons; XCaMP-Gf, 14 wells, 194 neurons; 2 independent transfections, four 96-well plates. The jGCaMP8 sensors were moderately linear and exhibited a large slope in the 1–10 AP range ($0.59 \leq R^2 \leq 0.82$; $0.18 \leq m \leq 0.28$), but less linear and exhibited a lower slope in the 1–40 AP range ($0.43 \leq R^2 \leq 0.75$; $0.052 \leq m \leq 0.081$). On the other hand, GCaMP6s, jGCaMP7c, and XCaMP-Gf better maintained their linearity throughout the 1-40 AP range, but they had generally lower slopes in the 1–10 AP range ($0.16 \leq m \leq 0.18$).

Extended Data Fig. 7 Sensor diffusion in cultured neurons studied with fluorescence recovery after photobleaching (FRAP).

a. Top, images of a representative cultured neuron expressing jGCaMP8m before (left) and immediately after (right) laser illumination. Asterisk indicates bleached region. Bottom, representative single-trial FRAP curves for jGCaMP8s (blue), cytoplasmic mEmerald (mEm-Cyto; pink) and EGFP-β-actin (green), normalized to pre-stimulation fluorescence values

and aligned to the FRAP laser pulse (yellow). Boxed area denotes zoomed-in region shown in b. *n* values indicate number of neurons tested in each condition for subsequent panels. Scale bar, 10 µm. b. Recovery curves of all tested variants (mean ± std.dev.). For clarity, only every 10th point in the trace is plotted. The color scheme is the same as in a – this panel also shows GCaMP6s (grey) and jGCaMP8m (dark red). c. Resistant fractions. The resistant fractions of GCaMP6s (0.3 ± 1.2%), jGCaMP8m (1.3 ± 0.5%), and jGCaMP8s (0.4 ± 0.7%) were not significantly different from a cytosolic GFP marker (mEm-Cyto, 0.9 ± 0.7%), but were significantly different from actin-bound GFP (EGFP-β-actin, 16.1 ± 11.4%; Welch's ANOVA with Dunnett's T3 multiple comparisons test; n.s.: $P > 0.45$). P values: GCaMP6s vs. jGCaMP8m, 0.46; GCaMP6s vs. jGCaMP8s, >0.9999; GCaMP6s vs. mEm-Cyto, 0.95; GCaMP6s vs. EGFP.B-actin, 0.0014; jGCaMP8m vs. jGCaMP8s, 0.46; jGCaMP8m vs. mEm-Cyto, 0.86; jGCaMP8m vs. EGFP.B-actin, 0.0030; jGCaMP8s vs. mEm-Cyto, 0.97; jGCaMP8s vs. EGFP.B-actin, 0.0019; *n* values same as in panel a. d. Recovery curves (mean ± std.dev.) of jGCaMP8m, jGCaMP8s and GCaMP6s, without ("reg") or with ("iono") added ionomycin to saturate sensor with Ca^{2+} (**Methods**). *n* values correspond to the number of neurons tested in each condition. Insets: percent resistant fraction. Box-whisker plots indicate the median and 25th–75th percentile range; whiskers indicate the shorter of 1.5 times the inter-quartile range or the extreme data point.

Extended Data Fig. 8 jGCaMP8 sensor characterization in adult *Drosophila* L2 visual system assay.

a. Responses of jGCaMP8f, jGCaMP8s, and XCaMP-Gf to the visual stimulus, as in Fig. 2b. b. Raw fluorescence intensity from the five sensors tested. Inset below: XCaMP-Gf shown with y-axis ~30x smaller. c. Mean intensity over the 0.5 Hz stimulation period shown in b. Kruskal-Wallis test finds P = 5.7e-5 and pairwise Dunn's multiple comparison test to jGCaMP7f as follows: jGCaMP8f = 6.5e-5, jGCaMP8m = 1.4e-2, jGCaMP8s = 0.37, and XCaMP-Gf = 2.8e-5; total *n* for each variant: j7f, 14 flies; jGCaMP8f, 14; jGCaMP8m, 11; jGCaMP8s, 11; XCaMP-Gf, 4. Bottom, images of mean intensity projection over the 0.5 Hz stimulation period, with color scale constant between variants. Scale bar, 5 µm. The jGCaMP8 indicators were dimmer than jGCaMP7f. d. Spectral power density measured from L2

responses at stimulation frequencies ranging from 0.5 to 30 Hz. e. $\Delta F/F_0$ responses to dark flashes 4, 8, or 25 ms in duration. Top, fluorescence traces show the mean ± std. dev. Bottom, box plots showing the sensitivity index d'. Kruskal-Wallis test followed by pairwise Dunn's multiple comparison test, *: $P < 0.05$. The shading in line plots in d and e represents standard error. In c and e, box-whisker plots indicate the median and 25th–75th percentile range; whiskers indicate the shorter of 1.5 times the inter-quartile range or the extreme data point. Complete statistics in Methods.

Extended Data Fig. 9 Expression of the GCaMP variants in adult fly visual system and larval neuromuscular junction.

a. Western blot analysis comparing protein expression between GCaMP variants. Ratio is the band intensity levels from a variant divided by the band intensity from the actin loading control. Multi-comparison Kruskal-Wallis finds P=0.038 and pairwise Dunn's multiple comparison test to jGCaMP7f as follows: jGCaMP8f = 0.011, jGCaMP8m = 0.019, jGCaMP8s = 0.024, and XCaMP = 0.038. Numbers tested are as follow: jGCaMP8f = 3, jGCaMP8m = 3, jGCaMP8s = 3, jGCaMP7f = 5, and XCaMP = 3. The jGCaMP8 and XCaMP variants expressed ~3x less protein than jGCaMP7f in L2 neurons. b. Box plot comparing immunostaining at the NMJ. Ratio is the intensity from stained variant divided by intensity from a myr::tdTomato co-expressed with the variant. Multi-comparison Kruskal-Wallis finds P = 0.029 and pairwise Dunn's multiple comparison test to jGCaMP7f as follows: jGCaMP8f = 0.37, jGCaMP8m = 0.039, and XCaMP = 4.2e-3. Numbers tested are as follow: jGCaMP8f = 2, jGCaMP8m = 6, jGCaMP7f = 3, and XCaMP = 2. c. Immunostaining females expressing GCaMP variants and myr::tdTomato in MBON-γ2α'1. Left, images from cell bodies (top), axons (middle), and dendrites (bottom). Scale bar is 1 µm. Green images show variant expression while red images show myr::tdTomato expression. Right, box plots quantify the ratio between intensity from the variant to the myr::tdTomato. Multi-comparison Kruskal-Wallis for cell body finds P = 0.05. Multi-comparison Kruskal-Wallis for axon finds P = 0.032 and P-values from pairwise Dunn's multiple comparison test as follows: jGCaMP8f = 0.13, jGCaMP8m = 0.018, and XCaMP = 0.010. Multi-comparison Kruskal-Wallis for dendrite finds P = 0.040 and p-values from pairwise Dunn's multiple comparison test as follows: jGCaMP8f =

0.079, jGCaMP8m = 0.034, and XCaMP = 0.010. Numbers tested are as follows: jGCaMP8f = 3, jGCaMP8m = 3, jGCaMP7f = 4, and XCaMP = 2. The jGCaMP8 variants expressed ~3x less protein than jGCaMP7f in L2 neurons. Box-whisker plots indicate the median and 25th–75th percentile range; whiskers indicate the shorter of 1.5 times the inter-quartile range or the extreme data point.

Extended Data Fig. 10 Characterization of GCaMP variants in larval neuromuscular junction (NMJ).

a. Design of larval NMJ experiments. b. Fluorescence response to 1, 5, 10, 20, 40, 80 and 160 Hz stimulation (2 s) of motor axons. Inset: zoomed response to 1, 5, 10 and 20 Hz. jGCaMP8s showed superior response from 1-20 Hz and jGCaMP7f above 80 Hz, where signals saturated. Mean ± s.e.m. shown. c. Saturating $\Delta F/F_0$ to 2 s motor axon stimulation at 1, 5, 10, 20, 40, 80 and 160 Hz. Mean ± s.e.m. shown. d. Half-rise time from stimulus onset to saturated peak under 40 Hz stimulation. Half-rise time at 40 Hz stimulation was markedly shorter than jGCaMP7f for all jGCaMP8 variants. e. Half-decay time from stimulus end to baseline under 40 Hz stimulation. Half-decay time was much shorter than jGCaMP7f for jGCaMP8f and jGCaMP8m. f. F_0 for each sensor. Dash line indicates the background fluorescence level. Resting fluorescence for the jGCaMP8 variants was lower than jGCaMP7f. g. Individual responses to 1, 5, and 10 Hz stimulation. The jGCaMP8 series detect individual stimuli much better than jGCaMP7f. Box-whisker plots in d-g indicate the median and 25^{th}–75^{th} percentile range; whiskers indicate the shorter of 1.5 times the inter-quartile range or the extreme data point. h. Power spectral density normalized to 0 Hz for responses to 1, 5, 10, and 20 Hz stimulation. Colors as above. Power spectral analysis confirms the performance of the jGCaMP8 indicators, with jGCaMP8m performing the best at all frequencies, particularly at the high end – jGCaMP8m shows strong power at 20 Hz trains, whereas jGCaMP7f is negligible. Panels d-g: Each data point represents a single bouton. # of boutons per line are: jGCaMP8f, 27; jGCaMP8m, 25; jGCaMP8s, 25; jGCaMP7f, 21. Boutons are from five individuals per line.

Extended Data Fig. 11 Responses across trials and long-term incubation.

a. Stable responses across trials. The peak response amplitude of orientation selective neurons was averaged and normalized (8f, 288 neurons; jGCaMP8m, 305 neurons; jGCaMP8s, 420 neurons; jGCaMP7f, 269 neurons; XCaMP-Gf, 121 cells) and plotted as a function of trial number. No stimulus adaptation was evident (mean ± s.e.m.). b-e. Response comparison between 3 weeks and 8 weeks post-AAV infection. b. Top, schematic of the experiment. Bottom, image of V1 L2/3 cells expressing jGCaMP8f eight weeks post-AAV injection (left), and the same field of view color-coded according to the neurons' preferred orientation (hue) and response amplitude (brightness). This experiment was repeated independently with similar results in 9 FOVs from 2 mice. c. Example traces from two L2/3 neurons in b. Light traces: five individual trials; dark traces: mean. Eight grating motion directions are indicated by arrows and shown above traces. The preferred stimulus is the direction evoking the largest response. Polar plots indicate the preferred orientation or direction of the cells. OSI values displayed above each polar plot. d. Box-plot comparison of half-decay time (in seconds) for jGCaMP8f between data acquired at 3 weeks and 8 weeks post-AAV injection. 225 cells from 6 mice for 3 weeks' data ([min, Q1, Q2, Q3, max] = [0.33, 0.71, 0.79, 0.89, 1.00]); 50 cells from 2 mice for 8 weeks' data ([min, Q1, Q2, Q3, max] = [0.33, 0.71, 0.79, 0.89, 1.00]). Two-sided Wilcoxon rank-sum test, $P = 0.60$. e. Comparison of peak response ($\Delta F/F_0$, %) for jGCaMP8f between data acquired at 3 weeks and 8 weeks post-AAV injection. 225 cells from 6 mice for 3 weeks' data ([min, Q1, Q2, Q3, max] = [12.0, 30.1, 42.9, 66.6, 396.4]); 50 cells from 2 mice for 8 weeks' data ([min, Q1, Q2, Q3, max] = [15.2, 26.4, 35.5, 65.1, 109.4]). Two-sided Wilcoxon rank-sum test, $P = 0.053$.

Extended Data Fig. 12 Sensor brightness *in vivo* and expression level.

a. Representative *in vivo* movie averages for all GECIs. The post-objective illumination power and the depth of imaging is noted under each image.

The brightness scale is the same for all images. b. *In vivo* distribution of excitation power-corrected baseline fluorescence values for segmented cellular ROIs. Horizontal bars represent the median of each distribution. Note the logarithmic scale. All data are normalized to the median of the jGCaMP7f distribution. See panel a for representative motion corrected *in vivo* two-photon movie averages. c. Representative images of anti-GFP fluorescence for all GECIs in a coronal section across the center of an injection site, 20–22 days post injection. The brightness scale is the same for all images. d. Distribution of somatic fluorescence values of anti-GFP antibody labelling for all sensors, 20–22 days post injection. Horizontal bars represent the median of each distribution. All data is normalized to the median of the jGCaMP7f values. Note that the expression levels are similar across sensors. The data is collected from two mice for each sensor. See panel c for representative images.

Extended Data Fig. 13 Orientation selectivity of the GCaMP-expressing mice.

a. Distribution of orientation selectivity index (OSI) for visually responsive cells measured using different sensors (n = 473 cells from 3 mice for jGCaMP7f; n = 221 cells from 3 mice for XCaMP-Gf; n = 484 cells from 5 mice for jGCaMP8f; n = 532 cells from 4 mice for jGCaMP8m; n = 742 cells from 5 mice for jGCaMP8s). There is a noticeable left shift in the distributions of OSI for jGCaMP8m and jGCaMP8s. b. Comparison of OSI values across sensors (same data as in a). jGCaMP7f ([min, Q1, Q2, Q3, max] = [0.010, 0.51, 0.71, 0.84, 1.0]); XCaMP-Gf ([min, Q1, Q2, Q3, max] = [0.0010, 0.38, 0.69, 0.83, 1.0]); jGCaMP8f ([min, Q1, Q2, Q3, max] = [0.0010, 0.48, 0.72, 0.85, 1.0]); jGCaMP8m ([min, Q1, Q2, Q3, max] = [0.012, 0.35, 0.57, 0.77, 1.0]); jGCaMP8s ([min, Q1, Q2, Q3, max] = [0.00030, 0.33, 0.53, 0.74, 1.0]). Kruskal-Wallis test ($P = 5.80 \times 10^{-26}$) with Dunn's multiple comparison test was used for statistics. jGCaMP7f *vs* XCaMP-Gf: $P = 0.13$; jGCaMP7f *vs* jGCaMP8f: $P = 1.0$; jGCaMP7f *vs* jGCaMP8m; $P = 1.1 \times 10^{-10}$; jGCaMP7f *vs* jGCaMP8s; $P = 2.0 \times 10^{-17}$; jGCaMP8m *vs* jGCaMP8s: $P = 1.0$. ***$P < 0.001$. ns, not significant.

Extended Data Fig. 14 Analysis of simultaneous imaging-electrophysiology experiments.

a-d. Descriptive statistics for loose-seal cell-attached recordings. a. Summary plot showing the number of mice used (bars, left y-axis) and the expression time at the time of the loose-seal recording in days (dots, right y-axis), for each sensor. b. Summary plot showing the total number of cells recorded (bars, left y-axis), and the number of cells recorded per mouse (dots, right y-axis) for each sensor. c. Summary plot showing the total length of simultaneous imaging and loose-seal recordings in hours (bars, left y-axis), and the length of simultaneous imaging and loose-seal recordings in minutes for each cell (dots, right y-axis). d. Summary plot showing the total number of action potentials (bars, left y-axis), and the number of recorded action potentials for each cell (dots, right y-axis – log scale), for each sensor. e-f. Signal-to-noise ratio of action potential recordings. e. Representative waveforms of loose-seal recorded action potentials in current-clamp (left) and voltage-clamp (right) recording mode. f. Signal-to-noise ratio distribution for all recorded action potentials in current-clamp (left) and voltage-clamp (right) recording mode. g-j. Sensor fluorescence across cell body ROIs and neuropil. g. A representative fluorescence trace for a cellular ROI (green) and its surrounding neuropil (blue) with simultaneous loose-seal recording. For calculating the distribution of neuropil contamination coefficients (r_*neu*), time points during the 3 s after an electrophysiologically recorded action potential (red vertical bars) were not included. Time points included in the analysis are highlighted in red. Note the correlation between cellular and neuropil ROI. Traces were high-pass filtered using a 10-second-long minimum filter and low-pass filtered with a Gaussian filter ($\sigma = 10$ ms). h. Cellular ROI pixel intensity values plotted against their corresponding neuropil pixel intensity values (time points highlighted with red in panel g), and their linear fit. The neuropil contamination coefficient is defined as the slope of this fitted function. i. Raw and neuropil corrected trace from panel g (40-80 sec), corrected with the neuropil contamination coefficient calculated in panel h (*F_corr = F_roi - r_neu*F_neu*). j. Distribution of *r_neu* values, each calculated on 3-minute-long simultaneous optical and electrophysiological recordings as shown in panels g-h. We included *r_neu* values only with a Pearson's correlation coefficient > 0.7. Colors represent different GECIs.

Calculated values of *r_neu* were similar between GECIs except for XCaMP-Gf, which was quite dim.

Extended Data Fig. 15 Effective ~500 Hz reconstruction of fluorescent responses *in vivo*.

a. Example isolated action potential during a simultaneous loose-seal recording at 50 kHz (top panel) and imaging at 122 Hz (bottom panel) of an jGCaMP8s-expressing neuron. b. Same as in a but 250 isolated action potentials are aligned to the peak of the action potential and overlaid. Note that frame times (green dots in middle panel) are uniformly distributed in time. Bottom, construction of the high-resolution resampled trace. Each point in the resampled trace is generated by averaging the surrounding time points across the population of calcium transients with a Gaussian kernel. Three example points are highlighted with black, red, and blue colors, together with the time span and weight used for the calculation of each point. c. Mean intensity projection of a representative field of view during cell-attached loose-seal recording. Recording pipette is highlighted with dashed white lines. The right panel shows how each frame is generated: the horizontal axis is scanned with a resonant scan mirror, the speed of which can be considered instantaneous relative to the vertical axis. The vertical axis is scanned with a slower galvanometer mirror, the speed of which determines the frame rate. d. Cellular ROI of the loose-seal recorded cell in panel c. Color scale shows pixel weights for ROI extraction. Right: cumulative pixel weight over the generation of a frame. We defined the timespan of the ROI as the 5-95% time of the cumulative pixel weight function. The timespan of the ROI is denoted with a red two-headed arrow. e. All loose-seal recorded ROIs weights overlaid as in panel d. An ROI was defined from three-minute-long movies, so a single recorded cell can have multiple overlapping ROIs in this image. f. Distribution of 5-95% timespans of all recorded ROIs. The timespans of most ROIs are under two milliseconds – thus the upper bound of the temporal resolution is ~500 Hz.

Extended Data Fig. 16 Responses in fast-spiking interneurons.

a. Spike waveform parameters for each recorded cell; colors represent the expressed sensor, and the size of the circle represents average firing rate.

Peak-to-trough ratios larger than 10 are plotted as 10. We defined putative interneurons as cells occupying the lower left quadrant (short peak-to-trough time and low peak-to-trough amplitude ratio), borders highlighted with red dotted lines. b. Example average action potential waveforms of a putative fast-spiking cell (black) and a putative pyramidal cell (green). The corresponding cells are marked with asterisks in panel a. c. Average calcium transient waveform for a single action potential in putative interneurons for jGCaMP8f, jGCaMP8m, and jGCaMP8s. Resampling was done with a 20-ms-long mean filter. d. Simultaneous fluorescence dynamics and spikes in jGCaMP8f (top), jGCaMP8m (middle) and jGCaMP8s (bottom) expressing putative interneurons. Fluorescence traces were filtered with a Gaussian filter ($\sigma = 5$ ms). e. Zoomed-in view of bursts of action potentials from dotted rectangles in panel d (top, jGCaMP8f; middle, jGCaMP8m; bottom, jGCaMP8s).

Extended Data Fig. 17 Imaging dendritic spikes in cerebellar Purkinje neurons.

a. Experimental design. Purkinje neurons in cerebellar lobule VI were transduced with a GCaMP variant as in the sample widefield (top right) and 2P (bottom right) images. Dendritic tufts were monitored for complex spike-related activity using 2P microscopy under free-locomotion conditions. b. Sample traces from adjacent dendrites for each variant. c. Half-decay times (Kruskal-Wallis P = 8.66e-11; Dunn’s test P values: 6f to 7f = 0.98, 6f to 8s = 0.99, 6f to 8m = 9.41e-7, 6f to 8f = 1.7e-7, 7f to 8s = 1, 7f to 8m = 0.0041, 7f to 8f = 4.62e-4, 8s to 8m = 0.0021, 8s to 8f = 2.42e-4, 8m to 8f = 0.99). d. Normalized fluorescence traces from the average of 10 events nearest to the median values from each variant. e. Half-rise times (Kruskal-Wallis P = 4.03e-26; Dunn’s test P values: 6f to 7f = 0.0025, 6f to 8s = 1.03e-7, 6f to 8m = 1.81e-10, 6f to 8f = 1.57e-6, 7f to 8s = 0.65, 7f to 8m = 0.10, 7f to 8f = 0.49, 8s to 8m = 0.99, 8s to 8f = 1, 8m to 8f = 1). f. Distribution of $\Delta F/F_0$ responses to complex spikes (Kruskal-Wallis P = 2.99e-9; Dunn’s test P values: 6f to 7f = 0.0010, 6f to 8s = 0.013, 6f to 8m = 1.22e-5, 6f to 8f = 0.67, 7f to 8s = 0.99, 7f to 8m = 0.99, 7f to 8f = 3.89e-4, 8s to 8m = 0.83, 8s to 8f = 0.0027, 8m to 8f = 1.40e-5). For each variant, 2 mice were imaged with number of dendrites per variant as: GCaMP6f, n = 51; jGCaMP7f, n = 14; jGCaMP8s, n = 14; jGCaMP8m, n = 13; jGCaMP8f,

n = 9. In box plots, boxes indicate median and inter-quartile range (IQR) while whiskers extend to the extrema or 1.5*IQR + (−) q3 (q1) with outliers lying beyond those values.

Extended Data Fig. 18 Statistics of S2F fits in the different imaging conditions.

a-f. Statistics of S2F fits in the different imaging conditions (See Extended Data Table 6 for more details). Blue, jGCaMP8f; red, jGCaMP8m; dark gray, jGCaMP8s; green, jGCaMP7f; cyan, XCaMP-Gf. a. Boxplots of rise time constant, $\boldsymbol{\tau}_r$. Minima, 0^{th} percentile of data (0%); maxima, 100%; center, 50%; bounds of box, from 25% (lower quartile) to 75% (upper quartile); whiskers, 1.5 times the distance between upper and lower quartiles. Number of biologically independent cells collected in each condition is summarized in Extended Data Table 5. b. Boxplots of half-rise time derived from S2F fits. Minima, 0^{th} percentile of data (0%); maxima, 100%; center, 50%; bounds of box, from 25% (lower quartile) to 75% (upper quartile); whiskers, 1.5 times the distance between upper and lower quartiles. Number of biologically independent cells collected in each condition is summarized in Extended Data Table 5. c. Comparison between half-rise time derived from S2F fits (x-axis) with that measured by super-resolution patch data (y-axis); paired two-sample sign-rank tests; two-sided. Red dashed line is the identity line. d. Scatter plots of decay time constants. X-axis, the slow decay time constant, $\boldsymbol{\tau}_{d2}$; y-axis, the fast decay time constant, $\boldsymbol{\tau}_{d1}$; size of dots, the ratio r of the weight for fast decay time to that for the slow one. Number of biologically independent cells collected in each condition is summarized in Extended Data Table 5. e. Box-plots of half-decay time derived from S2F fits. Minima, 0^{th} percentile of data (0%); maxima, 100%; center, 50%; bounds of box, from 25% (lower quartile) to 75% (upper quartile); whiskers, 1.5 times the distance between upper and lower quartiles. Number of biologically independent cells collected in each condition is summarized in Extended Data Table 5. f. Comparison between half-decay time derived from S2F fits (x-axis) with that measured by super-resolution patch data (y-axis; see Fig. 4e for more details); paired two-sample signed rank tests; two-sided. Red dashed line is the identity line. g, h. $\Delta F/F_{Synth}$ simulated from the S2F models of different sensors.

Simulations are based on S2F fits from the biologically independent cells collected in each condition; the number of cells in each condition is summarized in Extended Data Table 5. g. Normalized synthetic calcium latent dynamics, *c(t)*; solid lines, mean; shaded area, s.e.m. h. Simulated peak nonlinearity, *i.e.*, synthetic fluorescence response to different numbers of action potentials. Error bars, s.e.m. across cells. i,j. Measures of linearity of each indicator. Two linear models are shown in i and j. The closer the response curves to 1 (black dashed line, the linear model), the more linear the indicator response is to the number of action potentials. The measure is based on S2F fits from the biologically independent cells collected in each condition; the number of cells in each condition is shown in Extended Data Table 5. i. Normalized peak nonlinearity, where the synthetic fluorescence, $\Delta F/\,F(nAP)$, is normalized as: $\frac{\Delta F/F({nAP})}{\Delta F/F(1{AP})\times n}$, where $\Delta F/\,F(1\,{AP})$ is the peak response to a single action potential, *n* is the number of action potentials. Error bars, s.e.m. across cells. j. Normalized peak nonlinearity, where the synthetic fluorescence, $\Delta F/\,F({nAP})$, is normalized as: $\frac{\Delta F/F({nAP})}{\bar{\Delta F/F({nAP})}}$, where $\bar{\Delta F/\,F({nAP})}$ is the linear fit of $\Delta F/\,F$ predicted by the number of action potentials *n*. Error bars, s.e.m. across cells. The linear region (normalized peak nonlinearity is at 1, one-sample Wilcoxon signed rank test, $p < .05$) for 8s is from 1 to 5 action potentials; that for 8m is from 1 to 6 action potentials; that for 8f is from 3 to 8 action potentials; that for 7f is from 3 to 5 action potentials; that for XCaMP-Gf is from 2 to 8 action potentials.

Extended Data Fig. 19 Statistics of F2S fits.

a-d. Pairwise comparisons of F2S performance under different imaging conditions. Pairwise comparisons (two-sample rank-sum tests; two-sided) of indicators in each performance measure in Fig. 5e–h. The heatmap presents the significance, *i.e.*, *p*-value. The top row (mean) shows the statistics of the average. a. Fluorescence dynamics (fits compared to raw fluorescence); b. Spiking (fits compared to ground-truth spiking dynamics); c. F-score (spike detectability) using a linear F2S model; d. Spike-timing error using a linear F2S model. e-i. Statistics of F2S fits using a nonlinear model. e. Example trace and fit of a cell using a nonlinear F2S model –

using the same conventions as Fig. 5d-bottom. Top, variance explained of fluorescence dynamics, 93%; bottom, variance explained of spiking, 13%. f-g. Performance of fitting activity profiles. Violin plots, lines from top to bottom: 75%, 50%, 25% of data, respectively. f. Fluorescence dynamics; g. Spiking. h. Spike detectability. i. Spike-timing error.

Extended Data Table 1 Biophysical properties of initial sensors with different calmodulin-binding peptides used in this study
Extended Data Table 2 Data collection and refinement statistics of jGCaMP8.410.80
Extended Data Table 3 Characterization of jGCaMP8 in purified protein and in dissociated neuronal culture
Extended Data Table 4 Comparison of sensitivity and kinetics of jGCaMP8 to XCaMP-G, -Gf, and -Gf0 sensors for 1AP field stimulation
Extended Data Table 5 Statistics of the degree of nonlinearity of sensors measured by the difference of variance explained by S2F sigmoid from linear model (mean ± std.dev.)
Extended Data Table 6 Statistics of S2F parameter fits (mean ± std.dev.)

Supplementary information

Supplementary Tables

This file contains Supplementary Tables 3 and 4. Supplementary Table 3: Adult *Drosophila* L2 assay protocol. Supplementary Table 4. Summary of changes to the field stimulation pipeline between jGCaMP7 publication and current manuscript. See separate Excel files for Supplementary Tables 1 and 2.

Reporting Summary

Peer Review File

Supplementary Table 1

Sensor variants tested with the field stimulation assay, normalized to GCaMP6s. $\Delta F/F_0$, half-rise time, time-to-peak, half-decay time, signal-to-noise ratio, and normalized F_0 given for each of 1, 3, 10, 160 AP. *P*-values computed by the Mann-Whitney U-test for each variant, compared to GCaMP6s. Nucleic acid and protein sequence of each variant, as well as detailed name (if any) shown. Separate file.

Supplementary Table 2

Data on recorded cells, including GECI, mouse, neuron # per each mouse, putative cell type (pyramidal, interneuron), recording length, recording mode, total # of recorded spikes, and median spike SNR. Separate file.

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Zhang, Y., Rózsa, M., Liang, Y. *et al.* Fast and sensitive GCaMP calcium indicators for imaging neural populations. *Nature* **615**, 884–891 (2023).

https://doi.org/10.1038/s41586-023-05828-9

- Received: 08 November 2021
- Accepted: 10 February 2023
- Published: 15 March 2023
- Issue Date: 30 March 2023
- DOI: https://doi.org/10.1038/s41586-023-05828-9

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

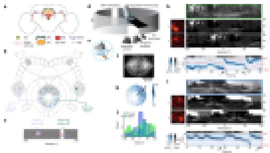
Provided by the Springer Nature SharedIt content-sharing initiative

Fastest-ever calcium sensors broaden the potential of neuronal imaging

- Michael B. Ryan
- Anne K. Churchland
- Casey Baker

Nature News & Views Forum 15 Mar 2023

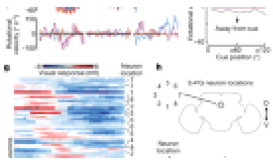
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05828-9



Generation of stable heading representations in diverse visual scenes

20 November 2019

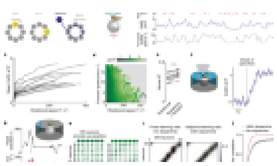
Sung Soo Kim, Ann M. Hermundstad, … Vivek Jayaraman



Sensorimotor experience remaps visual input to a heading-direction network

20 November 2019

Yvette E. Fisher, Jenny Lu, … Rachel I. Wilson



Dopamine promotes head direction plasticity during orienting movements

30 November 2022

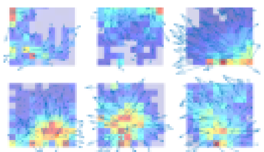
Yvette E. Fisher, Michael Marquis, … Rachel I. Wilson



[Flexible cue anchoring strategies enable stable head direction coding in both sighted and blind animals]

19 September 2022

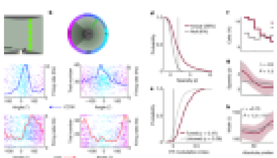
Kadjita Asumbisa, Adrien Peyrache & Stuart Trenholm



[Heading direction with respect to a reference point modulates place-cell activity]

27 May 2019

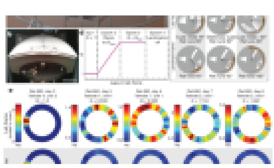
P. E. Jercog, Y. Ahmadian, … E. R. Kandel



[Moving bar of light evokes vectorial spatial selectivity in the immobile rat hippocampus]

09 February 2022

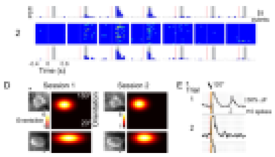
Chinmay S. Purandare, Shonali Dhingra, … Mayank R. Mehta



[Recalibration of path integration in hippocampal place cells](#)

11 February 2019

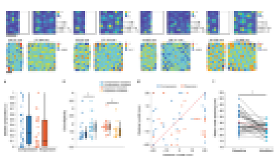
Ravikrishnan P. Jayakumar, Manu S. Madhav, … James J. Knierim



[Feature selectivity is stable in primary visual cortex across a range of spatial frequencies](#)

16 October 2018

Brian B. Jeon, Alex D. Swain, … Sandra J. Kuhlman



[Entorhinal velocity signals reflect environmental geometry](#)

13 January 2020

Robert G. K. Munn, Caitlin S. Mallory, … Lisa M. Giocomo

- Article
- Open Access
- Published: 22 March 2023

Population dynamics of head-direction neurons during drift and reorientation

- Zaki Ajabi ,
- Alexandra T. Keinath,
- Xue-Xin Wei &
- …
- Mark P. Brandon

Show authors

Nature **volume 615**, pages 892–899 (2023)

- 7568 Accesses

- 256 Altmetric

- Metrics details

Abstract

The head direction (HD) system functions as the brain's internal compass[1,2], classically formalized as a one-dimensional ring attractor network[3,4]. In contrast to a globally consistent magnetic compass, the HD system does not have a universal reference frame. Instead, it anchors to local cues, maintaining a stable offset when cues rotate[5,6,7,8] and drifting in the absence of referents[5,8,9,10]. However, questions about the mechanisms that underlie anchoring and drift remain unresolved and are best addressed at the population level. For example, the extent to which the one-dimensional description of population activity holds under conditions of reorientation and drift is unclear. Here we performed population recordings of thalamic HD cells using calcium imaging during controlled rotations of a visual landmark. Across experiments, population activity varied along a second dimension, which we refer to as network gain, especially under circumstances of cue conflict and ambiguity. Activity along this dimension predicted realignment and drift dynamics, including the speed of network realignment. In the dark, network gain maintained a 'memory trace' of the previously displayed landmark. Further experiments demonstrated that the HD network returned to its baseline orientation after brief, but not longer, exposures to a rotated cue. This experience dependence suggests that memory of previous

associations between HD neurons and allocentric cues is maintained and influences the internal HD representation. Building on these results, we show that continuous rotation of a visual landmark induced rotation of the HD representation that persisted in darkness, demonstrating experience-dependent recalibration of the HD system. Finally, we propose a computational model to formalize how the neural compass flexibly adapts to changing environmental cues to maintain a reliable representation of HD. These results challenge classical one-dimensional interpretations of the HD system and provide insights into the interactions between this system and the cues to which it anchors.

Main

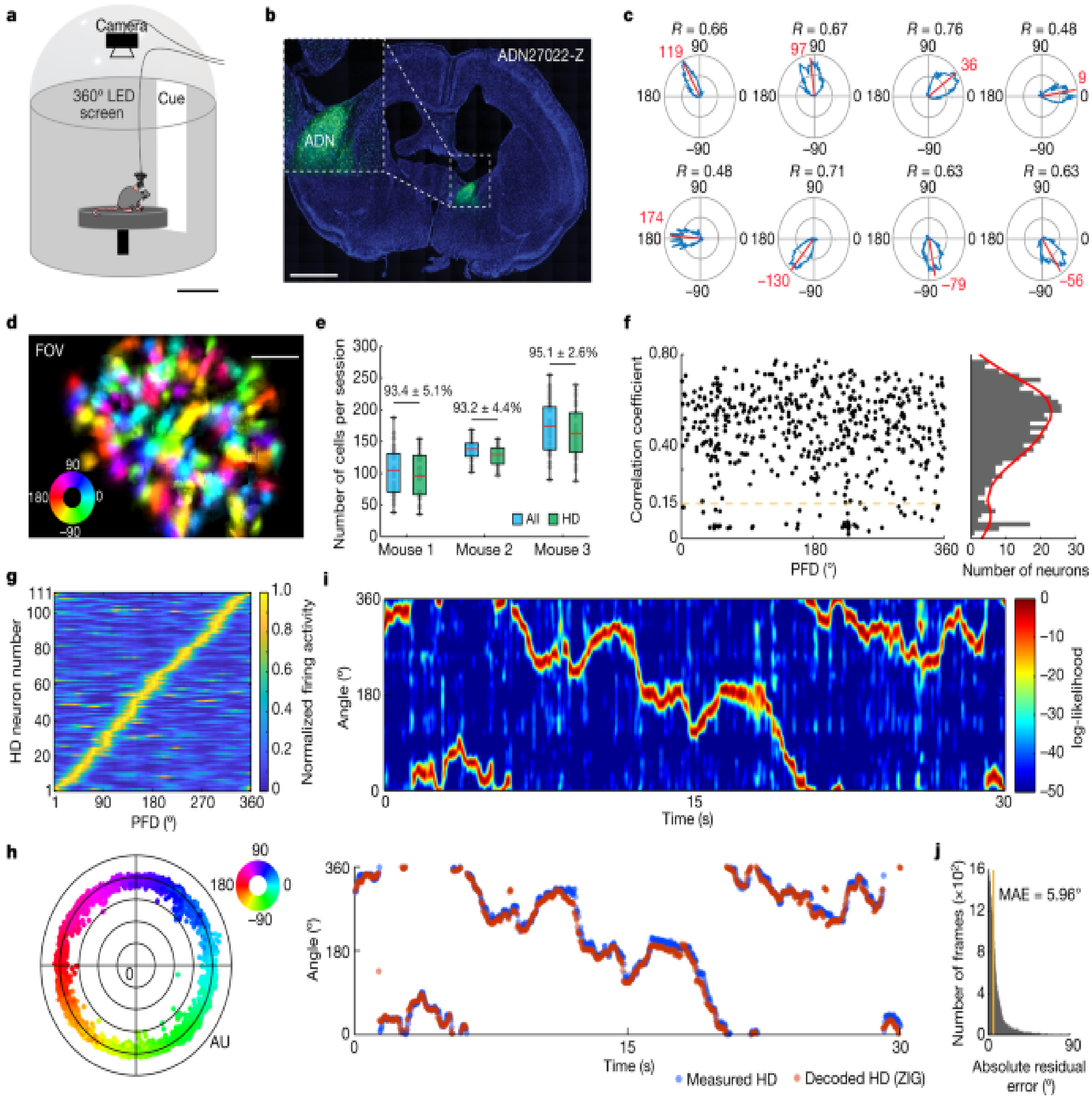
The HD system, commonly referred to as the neural compass, underlies a navigator's sense of direction[1,2,11,12,13,14]. In contrast to a traditional compass, the orientation of the HD system is anchored to local environmental cues[8,15,16]. Our understanding of the mechanisms that support the ability of the HD network to align with specific cues and maintain a consistent sense of direction remains limited. Recent research has shown that substantial variability in HD cell activity during sleep cannot be explained by a singular angular dimension[17,18]. This observation challenges the classical view of the internal HD representation as a unidimensional (that is, angular) construct and motivates further investigation of its complexity, from a population perspective, to understand whether extra dimensions are needed to fully capture how the system adapts to unstable conditions (that is, changing, missing and/or conflicting sensory information) in wakefulness. While updating the internal HD representation requires integration of information from multiple sensory modalities, manipulations of visual cues alone are sufficient to reorient this representation[6,7,19,20,21]. The visual input exerts a dominant influence on the HD network alignment, probably through a feedback correction that calibrates the integration of angular movements[22]. Computational models of the HD network suggest that plasticity mediates the integration of visual information within the network[23,24,25], confirmed recently in fruit flies[26,27] but not yet in mammals. Here we characterize the thalamic HD network response to visual manipulations, yielding new insights into the mechanisms that underlie anchoring and calibration of this

representation. Furthermore, we provide a network model to assess the plausibility of synaptic plasticity as a mechanism to explain the observed variability in the system's response to externally controlled changes in visual information.

Calcium imaging of the HD network

We performed calcium imaging of the anterodorsal thalamic nucleus (ADN) in three mice using a head-mounted endoscope[28,29,30]. Our recordings (n = 102 20 min sessions, with baseline periods of 3, 5 or 10 min depending on the experiment) enabled us to monitor up to 255 ADN cells simultaneously as mice freely explored a small elevated circular platform inside a larger enclosed chamber (Fig. 1a–g and Extended Data Fig. 1a–c). This chamber was composed of a 360° circular LED screen covered by an opaque dome (Fig. 1a). During baseline recordings at the start of each session, we displayed a polarizing vertical white stripe on the otherwise black LED screen. All subsequent testing involved the manipulation of this visual cue.

Fig. 1: Population recordings in the mouse ADN.



a, Schematic of the recording environment within a 360° LED screen. Scale bar, 20 cm. **b**, GCaMP6f expression in the ADN. In total, 12 mice were injected and implanted for this study, and only 3 (Extended Data Fig. 1a–c) provided enough simultaneously recorded HD cells for continued experimentation. Scale bar, 2 mm. **c**, Example tuning curves of ADN cells with high directional tuning in polar coordinates. The red lines and numbers show the mean resultant vectors and PFD, respectively. *R*, correlation coefficient. **d**, Field of view (FOV) of the ADN showing the PFD of each cell. Scale bar, 0.125 mm. **e**, The distribution of ADN cells recorded across mice ($n = 3$) and sessions ($n = 99$). The red line indicates the median (minimum, maximum, median, 25th percentile and 75th percentile,

respectively, are as follows: mouse 1 (all): 38, 188, 105, 70 and 131; mouse 1 (HD): 35, 154, 96, 66 and 128; mouse 2 (all): 102, 168, 138, 126.5 and 147; mouse 2 (HD): 97, 154, 129, 114.75 and 139.75; mouse 3 (all): 90, 255, 174, 137 and 204.5; mouse 3 (HD): 88, 239, 162.5, 133 and 195.5). The values above the box plots indicate the percentage of HD cells (green) among all recorded ADN cells (blue) shown as mean ± s.d. **f**, The distribution of correlation coefficients of ADN cells. The dashed yellow line represents the HD neuron detection threshold (shuffled control: $P < 0.05$). Data from three 10 min baseline recording sessions (one per mouse). **g**, HD population coverage of the azimuthal plane from one session. **h**, Projection of high-dimensional neural data onto a 2D polar plane using a feedforward neural network during a baseline recording. **i**, HD decoding. Top, log-likelihood distribution across time. Bottom, measured HD (blue) and decoded HD (red) using maximum likelihood. **j**, The distribution of the absolute residual error across baseline recordings from the first experiment ($n = 42$ sessions).

Calcium imaging data were motion-corrected and spiking activity was inferred from extracted fluorescent transients[31,32] (Extended Data Fig. 1d). Baseline recordings revealed HD cells with preferred firing directions (PFDs) that tiled the full 360° horizontal plane (Fig. 1g and Extended Data Fig. 1c). Consistent with previous research, ADN neurons were tuned to specific HDs[1,33,34] albeit with higher proportions (Fig. 1c–g and Extended Data Figs. 1 and 2). HD tuning was stable in the presence of visual cues[34] (Fig. 1c and Extended Data Figs. 1d and 2), and exhibited anticipatory firing[35] (Extended Data Fig. 3a). In contrast to the HD system in the central complex of *Drosophila*[36], we did not observe topographic organization (Fig. 1d and Extended Data Fig. 1b).

To infer the internal HD representation (referred to here as internal HD), we trained a Bayesian decoder[37] to estimate the animal’s HD on the basis of the baseline training data (Fig. 1i). This decoder accurately recovered the measured HD in stable experimental conditions (median absolute error (MAE) of test data = 5.96°; Fig. 1j).

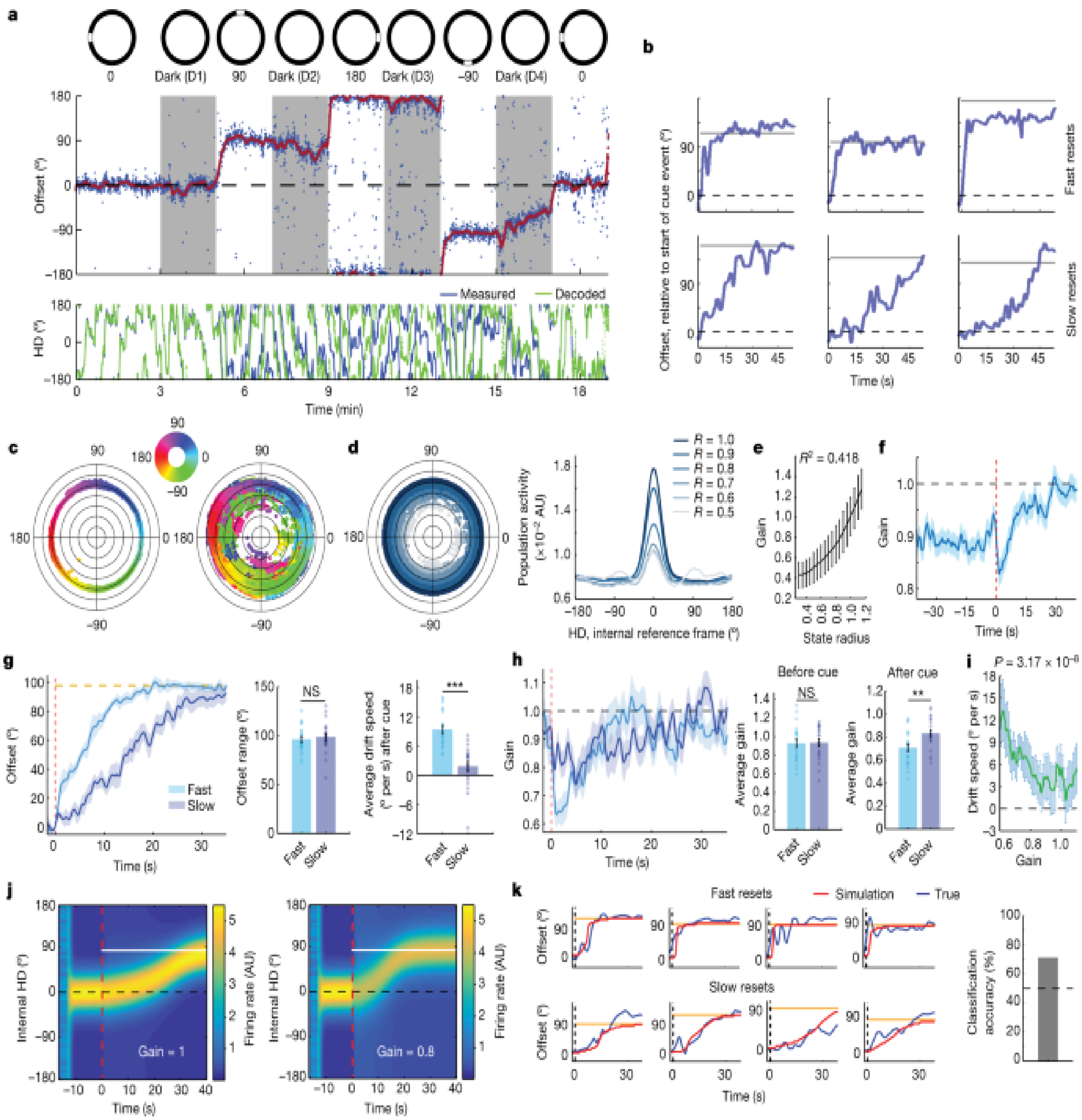
To visualize the low-dimensional structure of the HD representation, we developed a method to project large ensemble recordings onto a two-

dimensional (2D) polar state space (Methods). We trained a deep neural network on the measured head direction while allowing an untrained latent variable to capture variability in the neural data that cannot be explained by changes in the head direction alone. This inferred latent variable constitutes the radial component in the 2D polar state space (that is, secondary dimension). When applied to the baseline data, we obtained a ring-like structure (Fig. 1h), reminiscent of ring attractor models and previous analyses[3,17,38]. This further confirms that the internal HD representation is approximately unidimensional in stable conditions.

Network gain covaries with reset dynamics

To investigate how HD network dynamics enable reorientation, we recorded the HD network during a cue-shift paradigm. After a baseline recording, the cue was removed for 2 min (darkness) and then reappeared at a 90° shifted position for 2 min. We repeated this sequence four times per recording session (Fig. 2a).

Fig. 2: Network gain covaries with resetting dynamics.



a, Experimental protocol (top). Middle, example session showing the offset obtained by subtracting decoded from measured HD (blue dots). Red, smoothed offset. Darkness periods are shaded in grey. Bottom, measured (blue) and decoded (green) HD. **b**, Example fast (top) and slow (bottom) resets. The horizontal solid line indicates the cue location. The offset is relative to its angle at the cue onset. **c**, Projection of population activity onto the polar plane for the baseline (left) and the entire session (right). **d**, The same as in **c**; however, points are shaded by their radius (left). Right, mean bump of activity in the internal reference frame across radius ranges. **e**, The relationship between network gain and state radius ($n = 42 \times L$ datapoints,

where L is the number of frames in a session). R^2 value of linear regression model fit. Data are mean ± s.d. **f**, Triggered average of network gain (n = 168 = 4 × 42 cue events). The dashed red line indicates the cue display. **g**, Mean offsets for fast (light blue; n = 22 resets) and slow (dark blue; n = 20 resets) resets. Both groups have similar ranges (two-sided Wilcoxon rank-sum test, P = 0.4131, Z = 0.82), yet their speeds are different (two-sided Wilcoxon rank-sum test, $P = 1.0982 \times 10^{-6}$, Z = 4.87; 150 frames (~5 s) after the cue). **h**, Network gains of fast and slow reset groups have similar amplitudes before cue display (two-sided Wilcoxon rank-sum test, P = 0.6234, Z = 0.49; 50 frames (~1.67 s) before the cue), yet are different after cue display (two-sided Wilcoxon rank-sum test, P = 0.0085, Z = 2.63; 150 frames (~5 s) after the cue). The same data as in **g**. **i**, The relationship between gain and reset speed within 150 frames (~5 s) after the cue (n = 42 × 150 datapoints). The P value was calculated using an F-test on a linear model fit. **j**, Simulation of the bump of activity showing gain control of reset speed. The gain remains constant after cue display (dashed red). The solid white lines show the relative cue location. **k**, Model-based prediction (red) and true reset (blue). The dashed black lines indicate the cue display. The solid yellow lines indicate the relative cue location. All clockwise sessions were reflected across the x axis and transformed into counter-clockwise ones. Time-dependent signals in **f**–**h** are shown as mean ± s.e.m. Bar graphs and error bars, except in **e**, show mean ± s.e.m. with individual datapoints.

This manipulation resulted in predictable changes in the HD network's patterns of activity. To characterize these dynamics, we defined an 'offset' as the mismatch between the measured and decoded HD (see the 'Analysis of drift' section of the Methods). Tracking this offset, we observed a rotational response after cue reappearance that matched the cue shift. We refer to this phenomenon as a reset (Fig. 2a). Notably, resetting events were not homogeneous, as they occurred across a wide range of angles and speeds (Fig. 2b).

Cue shifts induced significant changes in the overall network activity. We observed modulation in the amplitude of the bump of activity (see the 'Reconstruction of the bump of activity' section of the Methods), which coincided with changes in the radius of the latent space (Fig. 2c,d). Intuitively, allowing the internal HD representation to occupy a 2D polar

state space makes the distance between any two given angles θ_1 and θ_2 a function of the radial component, which led us to hypothesize that changes in radius not only reflected changes in the overall population activity but would also correlate with changes in the speed of reset (Extended Data Fig. 4). To quantify this, we computed the total population activity normalized to the baseline activity, a measure that we refer to as the network gain (Extended Data Fig. 5; see the 'Calculation of network gain' section of the Methods). State-space radius was highly correlated with network gain (Fig. 2d,e), indicating that gain can be used as an interpretable measure of the radial component. To better understand the relationship between resetting events and network gain, we first analysed the 90°-centred reset range (that is, [70:110]° range). We found that the speed of HD-network reorientation, or 'reset speed', was anticorrelated with network gain (Fig. 2g–i). Separation of the resetting events into two groups (Fig. 2g) revealed that fast resets were associated with a substantial reduction in network gain shortly after cue reappearance, whereas slow resets exhibited a smaller reduction in gain (Fig. 2h and Extended Data Fig. 6a). In all cases, resetting events took the form of a continuous rotation of the HD representation from an initial orientation to the reset direction, passing by all intermediate angles without the appearance of secondary bumps of population activity (Extended Data Fig. 7a). This reset was slower than what has previously been reported[6], possibly due to differences in behaviour, habituation and/or the geometry of the testing set-up. Our modelling results show that an attractor network model that incorporated network gain replicated these dynamics with 71% accuracy in classifying fast versus slow resets (Fig. 2j,k and Extended Data Figs. 7b–d).

Behavioural differences as measured by the head angular velocity before and after cue onset could not explain the sharp decrease in gain amplitudes (Extended Data Fig. 8a). However, reduced head angular velocity immediately preceding cue events was predictive of fast resets, and vice versa (Extended Data Fig. 8b,c).

Resetting events also varied in the angular difference between their initial and stabilizing orientations. We grouped resets by the distance between pre-cue offset and the offset after stabilization (Extended Data Figs. 6b and 9a), and found that network gain was anticorrelated with reset range (Extended

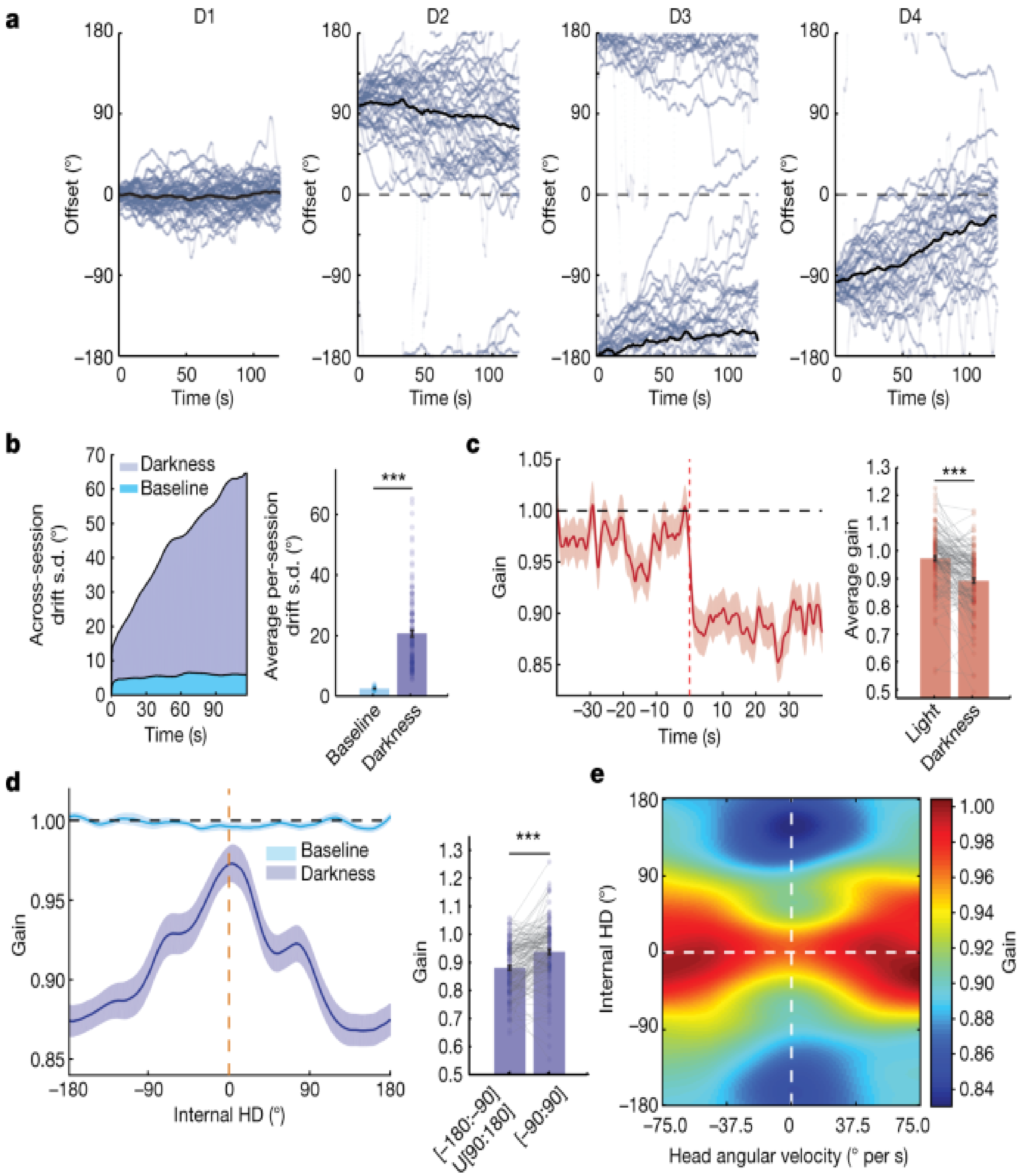
Data Figs. 6b and 9b,c). This relationship was independent of reset speed (Extended Data Fig. 9a), suggesting that network gain is independently modulated by the estimated error between the internal representation and the actual location of the visual cue.

We also detected a rapid spike in gain at the cue onset that was largest in short-range resets (Extended Data Fig. 9d). This may reflect visual inputs, but further investigation was limited by the temporal resolution of calcium imaging.

HD neurons maintain a trace of the cue

The HD network drifts in the absence of visual cues[5,8,9,10]. We hypothesized that network gain would decrease once the visual cue is removed due to decreased sensory input. This would bring the HD system to a lower energy state and cause the internal representation to become prone to spontaneous shifts because of the decrease in signal-to-noise ratio of neural activity. During all darkness epochs (D1 to D4), we observed an increase in drift relative to baseline (Fig. 3a,b and Extended Data Fig. 10a), which coincided with an abrupt decrease in the network gain after cue removal (Fig. 3c). Notably, changes in the network gain were dependent on the internal HD during darkness. When the internal HD pointed towards the internal location of the visual cue (0°), the reduction in gain was minimal; deviations from this direction resulted in more pronounced gain decreases (Fig. 3d,e). This gain profile persisted across all darkness periods with the difference between peak and trough increasing from the first to last darkness epoch (Extended Data Fig. 10d,e). Animal behaviour did not significantly affect the gain landscape, except that gain amplitude increased with the absolute head angular velocity (Fig. 3e and Extended Data Fig. 11), consistent with previous observations[36,39]. This suggests that the HD network maintains a ‘memory trace’ of salient visual cues.

Fig. 3: The network gain maintains a trace of the visual cue in darkness.



a, Offset traces (light blue) across darkness events D1 ($n = 42$), D2 ($n = 35$), D3 ($n = 33$) and D4 ($n = 35$). The black lines show the mean drifts. For D2, D3 and D4, only darkness epochs that follow a correct reset were considered. **b**, Drift variability increases with time during dark epochs compared with the baseline (baseline: $n = 42$ events; darkness: $n = 145$ events; two-sided Wilcoxon rank-sum test, $P = 7.5211 \times 10^{-23}$, $Z = 9.84$). **c**, The triggered average of network gain shows an abrupt drop at the transition

between cue-on and cue-off epochs, marked by the dotted red line (two-sided Wilcoxon rank-sum test, $P = 9.0165 \times 10^{-12}$, $Z = 6.81$; comparison between mean values over 40 s before and 40 s after cue removal; the same data as in **a**). **d**, Network gain tuning curves at the baseline (light blue) and during darkness (dark blue). The internal HD is relative to the baseline cue location (dashed yellow line). The gain remains flat at the baseline; however, it peaks around the internal cue location ([−90:90]°) and drops sharply away from it ([−180:−90]U[90:180]°) in darkness ($n = 145$ events; two-sided Wilcoxon rank-sum test, $P = 5.8683 \times 10^{-7}$, $Z = 5.00$). **e**, Network gain heat map. Note the increase in amplitude and width of the gain tuning curve at a larger head angular velocity. All clockwise sessions were reflected across the *x* axis and transformed into counter-clockwise ones. Signals in **c** and **d** are shown as mean (solid line) ± s.e.m. (shaded area). Bar graphs in **b**–**d** show the mean ± s.e.m. with individual datapoints.

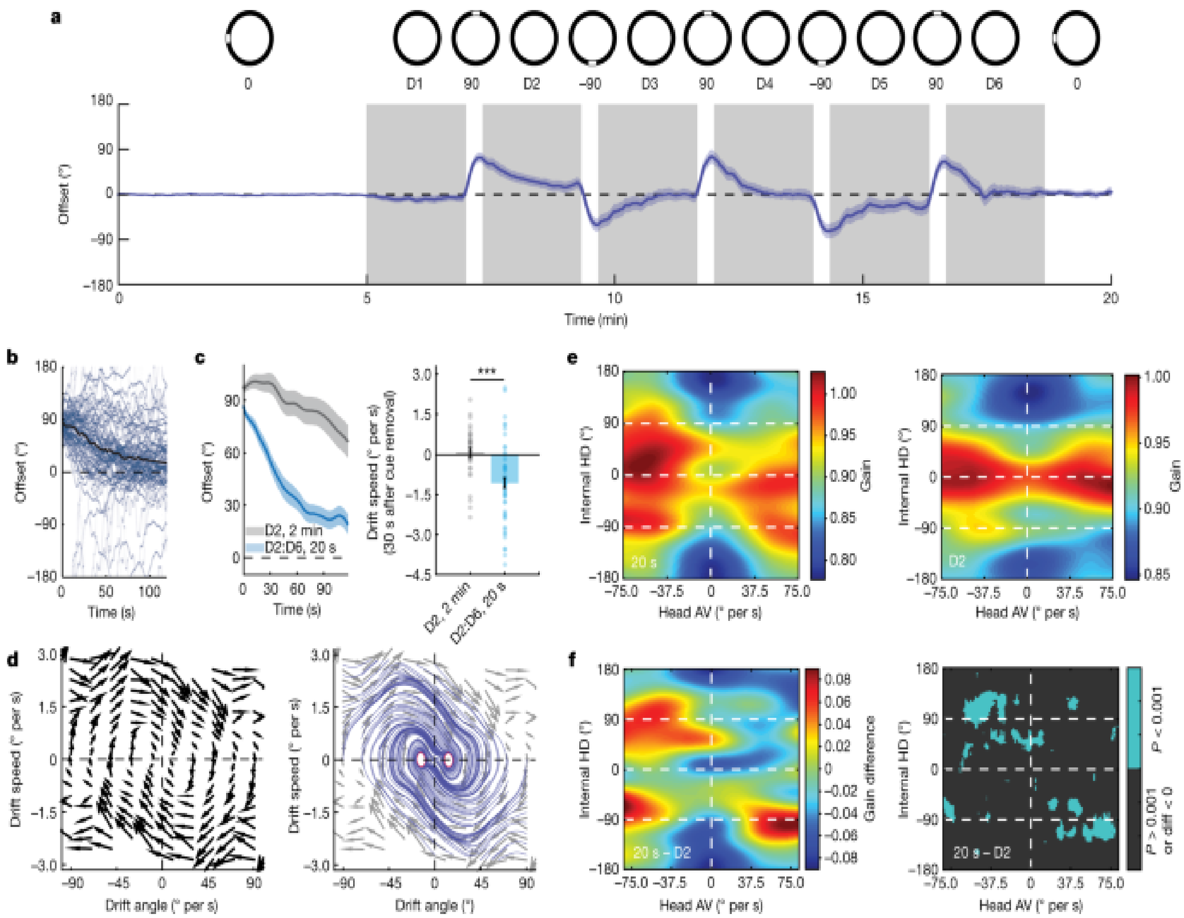
Drift patterns were not homogenous across the four darkness epochs (Extended Data Fig. 10a–c). During D1, drift fluctuated around the baseline orientation with no directional bias. By contrast, drift in D2, D3 and D4 exhibited directional biases dependent on the baseline orientation and previous cue location. During D2, drift diverged from its reset orientation towards its baseline orientation, counter to the rotation implied by the previous cue shift. During D3 and D4, drift rotated towards the baseline orientation but consistent with the direction implied by the previous cue shifts. These observations indicate that drift depends on previous visual experience. These predictable drift biases after exposure to the changing visual reference frame (D2 to D4) appeared to consistently bring the HD network closer to its original configuration (that is, the baseline state). Consecutive shifts of the visual cue in one direction further biased the drift in that direction (Extended Data Fig. 10b,c). These results suggest that both the stable allocentric and dynamic visual reference frames exert a persistent influence on the network orientation.

Drift patterns are experience-dependent

The presentation of a rotated visual cue for 2 min was sufficient to cause a representational shift and override the influence of non-visual cues (self-motion, olfactory and so on). Yet, we observed a tendency of the network to

rotate back towards the initial configuration, that is, revert to baseline, during darkness. We hypothesized that this could implicate plastic processes through which a ‘memory’ of baseline state exerts influence on the internal HD representation. If true, drift dynamics might depend on the duration of exposure to the shifted-cue context. To test this, we limited the display of the rotated visual cue to 20 s (±90° from the baseline; Fig. 4a). These shortened cue events elicited resets followed by reversions towards baseline during darkness (Fig. 4a–c). However, in comparison to the 2 min experiment (specifically D2, which was similarly preceded by a ±90° rotated cue event), reversion was much stronger after the presentation of a 20 s visual cue (Fig. 4c). The attraction of the network to its baseline state was further demonstrated through vector field analysis (Fig. 4d; see the ‘Vector field analysis’ section of the Methods). These results indicate that the internal representation of the baseline allocentric reference frame is not entirely lost after a reset and can still influence the HD network in darkness, depending on the duration of experience within the competing reset reference frame context. Addition of Hebbian learning to our model shows that, indeed, HD neurons could form new associations with the unchanged allocentric cues, depending on the duration of exposure to the reset context. Given enough time, the synaptic strength of these new associations increased while old associations were depleted, resulting in a new steady state. In this scenario, our simulations of the internal HD representation showed limited baseline attraction. By contrast, synaptic weights did not change significantly after shorter (20 s) cue exposures, and baseline associations remained dominant, resulting in strong reversions (Extended Data Fig. 12).

Fig. 4: Attraction of internal representation to the baseline reference frame is time dependent.



a, Experimental protocol (top row). Bottom row, average across-session offset (relative to baseline) during the 20 s cue-shift experiment. $n = 18$ sessions. Darkness periods are highlighted in grey. **b**, Individual offset traces after resets (light blue; $n = 58$ events) across darkness periods D2 to D6. The black line shows the mean offset. Offset traces after a −90° reset were reflected across the 0° axis. **c**, Mean offset during darkness in the 20 s cue-exposure experiment ($n = 58$ events; dark blue) and in D2 of the 2 min cue-exposure experiment ($n = 34$ events; grey) (left). Right, mean drift speed over the first 30 s shows a strong reversion to the baseline after a 20 s cue exposure (two-sided Wilcoxon rank-sum test, $P = 1.4264 \times 10^{-5}$, $Z = 4.34$). Data are mean ± s.e.m. with individual datapoints. **d**, Drift vector field (left). The arrows indicate the direction of mean drift speed and mean drift acceleration (from $n = 58$ events). The arrow length was scaled down for illustration. Right, simulated streamlines. The stable regime is highlighted in red. **e**, Network gain heat maps. Left, 20 s cue-exposure experiment. Data represent instances of reversion to the baseline ($n = 43$ events). Right, D2 of the 2 min cue-exposure experiment ($n = 34$ events). **f**, The gain difference

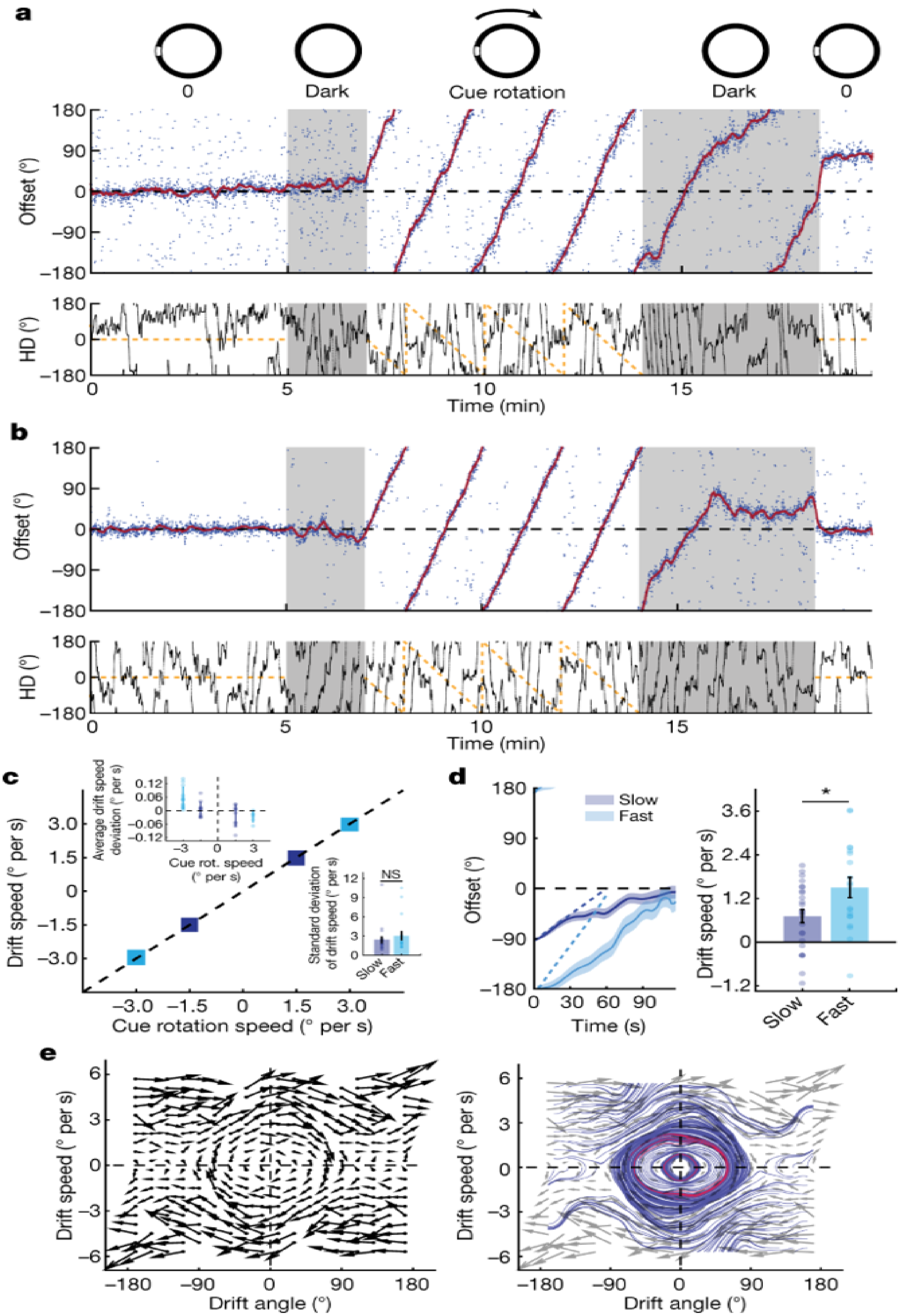
between heat maps in **e** showing the appearance of new bumps at the locations of cue-shifts (±90°) (left). Right, *P*-value matrix for the data on the left (two-sided Wilcoxon rank-sum test; pixels where $P > 0.001$ and/or gain (20 s) < gain (D2) are marked as not a number (NaN)). Time-dependent signals in **a** and **c** are shown as mean (solid line) ± s.e.m. (shaded area).

Next, we examined potential competition between baseline and reset reference frames by comparing network gain patterns after long and short shifted-cue exposures. While the gain landscape during D2 exhibited a single peak at 0° in the internal HD (corresponding to the shifted cue orientation after reset), the gain landscape during darkness after the 20 s cue events exhibited additional peaks at ±90° (Fig. 4e,f and Extended Data Fig. 13). Notably, these peaks matched the alternating ±90° cue structure of the experimental design, suggesting the coexistence of associations between HD neurons and allocentric cues from multiple visual contexts. These differences provide additional evidence of time-dependent effects of visual experience.

Cue rotation causes persistent drift bias

In the 2 min cue-shift experiment, visual information provided a dominant polarizing cue to reset the HD system. In some cases, resets were slow (>30 s), indicating that non-visual cues competed with visual information to stabilize the network. The 20 s cue-shift experiment provided further evidence that the baseline reference frame maintains a persistent influence on the HD network. To better understand the dynamics of this competition, we tested whether visual information could drive resetting when in continuous conflict with non-visual information, including self-motion cues. We recorded HD cell populations during presentation of a slow rotating visual cue (1.5 or 3.0° per s) for 7 min (Fig. 5a,b). In all cases, and for both speeds, the HD network was continuously updated by the visual cue (Fig. 5a–c), highlighting the dominant effect that the visual input has over all other inputs in controlling the HD system.

Fig. 5: Optic flow calibrates integration of change in HD.



a
0
Dark
Cue rotation
Dark
0
Offset (°)
180
90
0
−90
−180
HD (°)
180
0
−180
0
5
10
15
Time (min)
b
Offset (°)
180
90
0
−90
−180
HD (°)
180
0
−180
0
5
10
15
Time (min)
c
Drift speed (° per s)
3.0
1.5
0
−1.5
−3.0
−3.0
−1.5
0
1.5
3.0
Cue rotation speed (° per s)
Average drift speed deviation (° per s)
0.12
0.06
0
−0.06
−0.12
−3
0
3
Cue rot. speed (° per s)
Standard deviation of drift speed (° per s)
12
9
6
3
0
NS
Slow
Fast
d
Slow
Fast
Offset (°)
180
90
0
−90
−180
0
30
60
90
Time (s)
*
Drift speed (° per s)
3.6
2.4
1.2
0
−1.2
Slow
Fast
e
Drift speed (° per s)
6
3
0
−3
−6
−180
−90
0
90
180
Drift angle (°)
Drift speed (° per s)
6
3
0
−3
−6
−180
−90
0
90
180
Drift angle (°)

a, Experimental protocol (top). Middle, example offset (blue dots) during fast cue rotation (3° per s) showing persistent drift bias after cue removal. The solid red lines show low-pass filtered offset. Bottom, cue location (dashed yellow) and measured HD (solid black) relative to baseline cue location. Darkness periods are shown in grey. **b**, Example fast-cue-rotation session showing stabilization of the internal representation with an overshoot past the baseline orientation (top). Bottom, the same as in **a**. **c**, Mean drift speed during cue rotation (rot.) for fast (light blue) and slow (dark blue) sessions. Data are across-session mean ± s.d. Top left inset: subtraction of the cue-rotation speed from the average drift speed for individual sessions shows the average deviation of drift-speed with reference to the cue-rotation speed per session. Bottom right inset: comparison of drift-speed s.d. values between fast and slow sessions (two-sided Wilcoxon rank-sum test, $P = 0.5262$, $Z = 0.63$). **d**, The mean offset for fast (light blue; $n = 19$ events) and slow (dark blue; $n = 25$ events) sessions (left). The dotted lines correspond to the natural progression of offset if drift speed matched the speed of cue rotation. Data are mean (solid line) ± s.e.m. (shaded area). Right, drift-speed comparison between fast and slow sessions in the first minute after cue removal (Wilcoxon rank-sum test, $P = 0.0393$, $Z = 2.06$). All clockwise sessions were reflected across the x axis and transformed into counter-clockwise ones. **e**, Drift vector field (left). The arrows indicate the direction of mean drift speed and mean drift acceleration ($n = 60$ sessions). Arrow length was scaled down for illustration purposes. Right, simulated streamlines. The stable regime is highlighted in red. In **a**–**d**, fast sessions in which the offset angle at the beginning of the second darkness was within [−180:−145]U[145:180]° were considered, whereas slow sessions in which the offset's initial position in the second darkness was within [−125:−55]° were included. In **e**, all sessions were considered regardless of the offset angle at the start of the second darkness. For **c** and **d**, data are mean ± s.e.m. with individual datapoints.

Notably, the HD network continued to rotate in the same direction and at a similar speed when the rotating cue was turned off (Fig. 5a,b,d and Extended Data Fig. 14a; see the ‘Analysis of cue-rotation sessions’ section of the Methods). This persistent bias was replicated in our network model by adding a recalibration circuit to asymmetrically change the strength of vestibular input through visual feedback (Extended Data Fig. 14b–e and Supplementary Information). We also observed an attraction to the baseline

internal representation, similar to our prior experiments. The system started to stabilize once the internal HD representation approached the baseline state (Fig. 5e and Extended Data Fig. 14a). This phenomenon could also be reproduced in our model in which, after 7 min of cue rotation, no strong new associations between HD neurons and allocentric cues could emerge to form a new steady state. Instead, baseline associations remained dominant, albeit with a significant weight decay in the synaptic matrix (Extended Data Fig. 14d). Together, these results indicate that experience with dynamic reference frames can bias the HD network and implicate asymmetric recalibration of vestibular input integration within the HD network as a potential source of this bias.

Discussion

We combined large population recordings of ADN neurons with visual cue manipulations to characterize the population dynamics of the mammalian HD system. Controlled manipulations of a visual cue induced global fluctuations in network activity captured by a measure that we termed network gain. Network gain represents a functional dimension in the internal HD representation, in addition to its classically appreciated angular dimension. Reorientations, in response to visual cue shifts, were associated with a transient decrease in network gain, the magnitude of which was correlated with the speed of the HD network's realignment with the rotated visual reference frame (that is, reset). By extending a standard model of the HD system[40] to incorporate network gain, we were able to predict the speed of the reset response. These results suggest that modulation of network gain provides the HD system with a mechanism to rapidly reorient. Network gain also reflected the past experience of the system—a polarizing visual landmark induced persistent distortions in the network gain profile, forming a memory trace in darkness. These network gain patterns were dependent on the duration of the previous shifted-cue exposure, suggestive of plastic processes in the HD network. Evidence for plasticity was further strengthened by experience-dependent drift behaviours in darkness periods. Incorporating Hebbian plasticity into our network model replicated these observations. Finally, the HD system anchored to a continuously rotating visual cue and continued to rotate after the cue was removed. A model of asymmetric vestibular input recalibration reproduced these results,

suggesting that the integration of vestibular information within the HD network is also experience dependent.

Network gain reduction during realignment of the HD network suggests that a feedback signal downstream of ADN provides global inhibition to the network. Similar ideas have been proposed in the central complex of fruit flies[41,42]. Modulation of global neural activity might allow the HD system to operate at different energy levels with varying degrees of stability, reflecting uncertainty in the HD representation. We hypothesize that the animal's engagement in exploratory behaviour together with increased familiarity with the experimental environment and its geometric specificities could sustain a high-gain/high-certainty regime of operation and cause resistance to HD network reorientations imposed by visual cue shifts.

Recent research in fruit flies demonstrated plasticity between visual input and the compass neurons[26,27]. The current study complements these studies and provides evidence for experience-dependent influence of visual landmarks in mammals. Indeed, the mammalian brain appears to maintain a memory of the associations between HD neurons and visual landmarks, in the form of preferential firing, long after landmarks disappear. We propose that memory traces of salient cues in ADN cells help to stabilize the HD system during navigation, even in the absence of reliable environmental anchors by maintaining high activity levels (that is, high signal-to-noise ratio) around internal cue locations. Whether these network gain bumps have an active role in guiding navigation behaviour remains an open question. Moreover, baseline-state attraction during darkness is further evidence of long-term effects in the HD system. As the strength of this attraction depends on the duration of exposure to the shifted-cue context, we speculate that the underlying mechanisms leading to such behaviour involve synaptic plasticity. Through network modelling, we demonstrated that, by adding Hebbian learning, the observed drift patterns could be replicated. Our model proposes that neurons akin to the ring cells of fruit flies could mediate experience-dependent drift dynamics. Whether such neurons exist in the mammalian brain is yet to be determined.

The fact that a continuously rotating visual scene caused persistent biases in the HD representation, as demonstrated in our cue-rotation experiment, suggests that the integration of vestibular inputs undergoes an experience-

dependent calibration. Our findings complement a similar finding in place cells[22] and support a model of hierarchical transfer of information from HD neurons to downstream cells of the navigation system (that is, place cells, grid cells and so on) to maintain consistent and flexible cognitive maps[1,41,43,44].

Ultimately, our findings provide insights into the mechanisms that govern realignment and stabilization of the HD network, and how long-term effects of previous experience affect its dynamics. Importantly, these findings highlight the complexity of the internal HD representation and motivate studying this cognitive system in a multidimensional framework. Here we show evidence for the functional importance of the global fluctuations in network activity (that is, gain) as a critical, yet previously underappreciated, dimension. Future studies examining the origins of such fluctuations will be critical to unveil the complete picture of the intrinsic structure of this circuit.

Methods

Mice

Twelve male wild-type mice (C57Bl/6, Charles River) were used for this study, three of which provided enough simultaneously recorded HD cells for continued experimentation. Mice were housed individually at 22 °C under a 12 h–12 h light–dark cycle and 40% humidity with food and water ad libitum. All of the experiments were performed in accordance with McGill University and Douglas Hospital Research Centre Animal Use and Care Committee (protocol 2015-7725) and in accordance with Canadian Institutes of Health Research guidelines.

Surgeries

During all surgeries, mice were anesthetized by inhalation of a combination of oxygen and 5% isoflurane before being transferred to the stereotaxic frame (David Kopf Instruments), where anaesthesia was maintained by inhalation of oxygen and 0.5–2.5% isoflurane for the duration of the surgery. Body temperature was maintained with a heating pad and eyes were hydrated with gel (Optixcare). Carprofen (10 ml kg^{-1}) and saline (0.5 ml)

were administered subcutaneously, respectively, at the beginning and end of each surgery. Preparation for recordings involved three surgeries per mouse. First, at the age of 7–8 weeks, each mouse was injected with 600 nl of the non-diluted viral vector AAV9.syn.GCaMP6f.WPRE.eYFP, sourced from University of Pennsylvania Vector Core. All injections were administered through glass pipettes connected to the Nanoject II (Drummond Scientific) injector at a flow rate of 23 nl s^{-1}. One week after injection, a 0.5-mm-diameter gradient refractive index (GRIN) relay lens (Go!Foton) was implanted above the ADN (AP, 1.8; ML, 0.8; DV, −3). No aspiration was required. In addition to the GRIN lens, three stainless steel screws were threaded into the skull to stabilize the implant. Dental cement (C&B Metabond) was applied to secure the GRIN lens and anchor screws to the skull. A silicone adhesive (Kwik-Sil, World Precision Instruments) was applied to protect the top surface of the GRIN lens until the next surgery. Then, 2 weeks after lens implantation, an aluminium baseplate was affixed by dental cement (C&B Metabond) to the skull of the mouse, which would later secure the miniaturized fluorescent endoscope (miniscope) in place during recording. The miniscope/baseplate was mounted to a stereotaxic arm for lowering above the implanted GRIN lens until the field of view contained visible cell segments, and dental cement was applied to affix the baseplate to the skull. A polyoxymethylene cap was affixed to the baseplate when the mice were not being recorded to protect the baseplate and lens. After surgery, animals were continuously monitored until they recovered. For the initial 3 days after surgery, the mice were provided with a soft diet supplemented with Carprofen for pain management (MediGel CPF). Screening and habituation to recording in the experimental environment began 2–3 days after the baseplate surgery. The first 3–4 weeks of recordings were used to confirm the quality and reliability of the calcium data while the animal was exploring the environment with different screen displays.

Data acquisition

In vivo calcium videos were recorded with a miniscope (v3; https://miniscope.org) containing a monochrome CMOS imaging sensor (MT9V032C12STM, ON Semiconductor) connected to a custom data acquisition (DAQ) box (https://miniscope.org) with a lightweight, flexible

coaxial cable. The cable was attached to a noiseless pulley system with a counterbalance (placed outside the recording environment) to prevent interference with the recorded animal's movements and to alleviate the weight of the miniscope. The DAQ was connected to a PC with a USB 3.0 SuperSpeed cable and controlled using Miniscope custom acquisition software (https://miniscope.org). The outgoing excitation LED was set to 3–6%, depending on the mouse, to maximize the signal quality with the minimum possible excitation light to mitigate the risk of photobleaching. The gain was adjusted to match the dynamic range of the recorded video to the fluctuations of the calcium signal for each recording to avoid saturation. Behavioural video data were recorded using a webcam mounted above the environment. The DAQ simultaneously acquired behavioural and cellular imaging streams at 30 Hz as uncompressed AVI files and all recorded frames were timestamped for post hoc alignment. Two controllable LEDs (green and red) were added and used for tracking such that, whenever the miniscope was attached to the baseplate, the green LED pointed to the right side of the mouse's head and the red LED pointed to the left side. All other light sources from the miniscope were covered. All recordings took place inside a 360° LED screen (height: 1 m, diameter: 90 cm; Shenzhen Apexls Optoelectronic), at the centre of which we placed a wall-less circular platform (diameter, 20 cm) raised 50 cm above the ground. Mouse bedding was evenly spread over the platform before each recording session. In all recordings, mice were free to move on top of the raised platform. A half spherical dome was used to cover the environment and prevent external light from entering, while it also held the behavioural camera. The experimental environment was designed to maximize circular symmetry, in the absence of any screen display. During habituation, mice were recorded while exposed to a single vertical stripe or no visual display (darkness). These recordings were also used to confirm the quality of tracking the head direction and the cue location, in different conditions. In all experiments in this study, the visual cue refers to a single white vertical stripe (width, 15 cm; height, 1 m).

Data preprocessing

Calcium imaging data were preprocessed before analyses through a pipeline of open-source MATLAB (MathWorks; v.R2015a) functions to correct for motion artifacts[45], segment cells and extract transients[32,46], and infer the

likelihood of spiking events by deconvolution of the transient trace through a first-order autoregressive model[31]. We wrote a MATLAB (MathWorks, v.2015a) program to perform offline tracking of the LEDs and determine, at each frame, the animal's head direction. Another custom-written program was used to estimate the location of the visual cue. Both scripts were incorporated into the preprocessing pipeline.

Data analysis

In this study, neural activity refers to the deconvolved calcium traces as described previously[31] unless specified. The resulting time series (per neuron, per session) correspond to the inferred likelihood of spiking events. A moving average filter of width 3 frames (~100 ms) is then applied on each time series. We refer to the obtained signal as firing activity.

Identification of HD cells

For every identified cell segment (ROI), we construct an HD tuning curve by measuring the occupancy-normalized firing activity within each angle bin (1° per bin) of the horizontal plane (*x* axis). The tuning curve is circularly smoothed with a moving average filter of width 50°. This enables us to have a better estimate of the angle bin that corresponds to the maximum firing activity of a given neuron's tuning curve, which we will refer to as the PFD. We next construct a stimulus signal for that specific PFD by convolving the measured HD signal (from the behavioural camera) with a narrow Gaussian kernel (mean = PFD, s.d. = 17°) such that for every neuron *i*:

$${{\rm{stim}}}_{i}\left(t\right)={{\rm{e}}}^{-\frac{{({\rm{angdiff}}({{\rm{PFD}}}_{i},{\theta }_{{\rm{HD}}}(t)))}^{2}}{2{\sigma }^{2}}}$$

Where, $\theta_{\rm HD}$ is the measured HD time series, σ is the s.d. of the Gaussian kernel and, angdiff (a, b) is a MATLAB function that gives the subtraction of a from b, wrapped on the [−π,π] interval. We correlate the stimulus signal with a normalized version of the firing activity to obtain the Pearson correlation coefficient r of each neuron. To determine the threshold value of r above which a cell can be identified as an HD neuron, we used data from

ten baseline recordings (3 min) per animal, randomly selected from the reset experiment. We start with a relatively high value $r_{\rm{thresh}}$ and select all neurons such that $r > r_{\rm{thresh}}$. For each neuron, we produce 1,000 shuffles of the firing activity using the MATLAB circshift function (to preserve the temporal correlation of the firing activity signal) at random shifts. We next correlate each shuffled version with the stimulus signal of the corresponding neuron to obtain a distribution of correlation coefficients (3 separate distributions, 1 per mouse). We define ${r}_{m}^{95{\rm{th}}}$ as the value that corresponds to the 95th percentiles of the distribution, for mouse *m*. If ${r}_{{\rm{thresh}}} > {r}_{m}^{95{\rm{th}}}$, we keep iterating the same procedure while decreasing $r_{\rm{thresh}}$ by 0.01 until convergence (that is, ${r}_{{\rm{thresh}}}\simeq {r}_{m}^{95{\rm{th}}}$), which constitutes the correlation coefficient threshold to identify HD neurons for mouse *m* (see Extended Data Fig. 1d,e for an illustration of the results).

HD decoding from neural data

We trained a recently developed Bayesian decoder[37] to infer the HD direction from the deconvolved calcium responses of the imaged neural population. Noise independence across neurons was assumed. Conceptually, this decoder is similar to the Bayesian decoding method for spike trains as commonly used in the literature[47], except that we used zero-inflated-gamma distribution to model the stochasticity of the deconvolved calcium responses, instead of Poisson distribution. Our previous results showed that the zero-inflated-gamma model could better capture the noise of the calcium signal and provide better decoding results compared with the Poisson noise model and a few other alternatives. Details of this procedure can be found in section 4 of ref. [37]. Here we smoothed the log-likelihood matrix (rows, angle bins; columns, frames) by iteratively summing the likelihoods over 5 frames (~166.7 ms) centred around the corresponding timestep of each iteration, for each angle bin. Note that, owing to the predominance of HD-tuned neurons among detected cell segments and to avoid selection biases, the neural activity from all ADN cells was used as an input to the decoding algorithm.

Analysis of drift

We define the offset as the angular difference between the measured head direction ($\theta_{\rm measured}$) and the decoded head direction ($\theta_{\rm decoded}$):

$${\rm{Offset}}(t)={\rm{angdiff}}({\theta }_{{\rm{decoded}}},{\theta }_{{\rm{measured}}})$$

In all analyses involving drift estimation, both measured and decoded HDs were smoothed with a moving average filter of width 20 frames (~667 ms). For the analysis of drift during darkness (except for heat maps), further smoothing was applied to extract the low-frequency component of the signal whereby a moving average filter of width 300 frames (~10 s) was used. In all cases, a simple linear regression was performed on the unwrapped offset signal over a sliding window of 20 frames (~667 ms) to estimate the drift speed at the centre of the regression window (that is, slope of the fitted line).

Separation of fast and slow resets

Classification of resets within the [70:110]° range was done using the *k*-means clustering function in MATLAB. We used data from the first 1,450 frames after cue display. The algorithm separates between two clusters by generating 50 replicates with different initial cluster centroid positions for each replicate and then calculating the sums of point-to-centroid distances for each cluster using the ‘city block’ distance metric.

Reconstruction of the bump of activity

At any given time, we can reconstruct the bump of activity from the firing activity of each neuron and their respective tuning curves using a normalized weighted sum of tuning curves[48]:

$$A\left(\theta ,t\right)=\frac{\sum _{i}{f}_{i}(\theta){r}_{i}(t)}{\sum _{i}{f}_{i}(\theta)},$$

where, A is a 360-by-T matrix (each row is a 1° bin of the horizontal plane and each column is a frame within range T of the analysis), f_i is the tuning curve of neuron i and r_i is the firing activity of neuron i.

Calculation of network gain

We assume that, at any given time, the thalamic HD network is subject to a global gain modulation of the firing activity, applied homogeneously on all ADN neurons such that:

$$\begin{array}{cc}{r}_{i,t}={\alpha }_{t}{f}_{i}\left({\theta }_{t}\right)+\varepsilon , & \varepsilon \sim {\mathscr{N}}(0,{\sigma }^{2})\end{array},$$

where $r_{i,t}$ is the instantaneous firing activity of ADN neuron i; α_t is the instantaneous gain factor; f_i is the tuning curve of ADN neuron i (calculated on the basis of the response measured in the baseline condition); θ_t is the decoded head direction from neural activity at time t; and ε is the additive Gaussian noise.

Our goal is to estimate the value of α_t at any given time t using maximum-likelihood estimation approach.

Given the decoded head direction at time t, θ_t as well as the tuning curves f_i for all ADN neurons, we obtain the likelihood of observing $r_{i,t}$ with parameter α_t:

$$P\left({r}_{i,t}| \,{f}_{i}\left({\theta }_{t}\right);{\alpha }_{t}\right)={\mathscr{N}}({\alpha }_{t}\,{f}_{i}\left({\theta }_{t}\right),{\sigma }^{2}).$$

We define the vectors:

$$\begin{array}{cc}{R}_{t}=\left[\begin{array}{c}{r}_{1,t}\\ {r}_{2,t}\\ \vdots \\ {r}_{N,t}\end{array}\right], & F({\theta }_{t})=\left[\begin{array}{c}{f}_{1}({\theta }_{t})\\ {f}_{2}({\theta }_{t})\\ \vdots \\ {f}_{N}({\theta }_{t})\end{array}\right]\end{array},$$

where N is the number of ADN neurons in the network.

Assuming independent activity between said neurons, we can calculate the likelihood of observing R_t:

$$\begin{array}{l}P({R}_{t}| F({\theta }_{t})\,;{\alpha }_{t})={\prod }_{i}P({r}_{i,t}| \,{f}_{i}({\theta }_{t});{\alpha }_{t})\\ \,\,\,\,\propto {\prod }_{i}{\rm{\exp }}\left(\frac{{({r}_{i,t}-{\alpha }_{t}{f}_{i}({\theta }_{t}))}^{2}}{-2{\sigma }^{2}}\right).\end{array}$$

We apply the logarithm on both sides:

$$\log \left(P\left({R}_{t}| F\left({\theta }_{t}\right)\,{\rm{;}}{\alpha }_{t}\right)\right)\propto \,-\,\frac{1}{2}{\sum }_{i}\frac{{({r}_{i,t}-{\alpha }_{t}{f}_{i}\left({\theta }_{t}\right))}^{2}}{{\sigma }^{2}}.$$

Our goal is to determine the parameter ${\hat{\alpha }}_{t}$ that maximizes the log-likelihood such that:

$${\hat{\alpha }}_{t}=\mathop{{\rm{a}}{\rm{r}}{\rm{g}}{\rm{m}}{\rm{a}}{\rm{x}}}\limits_{{\alpha }_{t}}\,\log ((P({R}_{t}|F({\theta }_{t})\,;\,{\alpha }_{t}))=\mathop{{\rm{a}}{\rm{r}}{\rm{g}}{\rm{m}}{\rm{i}}{\rm{n}}}\limits_{{\alpha }_{t}}{\sum }_{i}{({r}_{i,t}-{\alpha }_{t}{f}_{i}({\theta }_{t}))}^{2}.$$

To do so, we take the derivative of the objective function with reference to α_t and set it to zero:

$$\frac{d}{d{\alpha }_{t}}{\sum }_{i}{({r}_{i,t}-{\alpha }_{t}{f}_{i}({\theta }_{t}))}^{2}=0.$$

Thus:

$${\hat{\alpha }}_{t}=\frac{\sum _{i}{r}_{i,t}\,{f}_{i}\left({\theta }_{t}\right)}{\sum _{i}{{f}_{i}\left({\theta }_{t}\right)}^{2}}.$$

Similar to the offset signal, the obtained gain is smoothed with a moving-average filter of width 20 frames (~667 ms), unless otherwise specified. With the exception of Extended Data Fig. 5b–d, the gain was always normalized to the average baseline activity. In Extended Data Fig. 5b–d, the

tuning curves represent the average firing rates as a function of the internal HD for the entire recording session.

Gain heat-map analysis

Gain heat maps are 2D matrices in which each pixel $p(x, y)$ is a 2D bin of width 1.5° per s corresponding to the measured angular head velocity and height 1° corresponding to the decoded HD. Pixel $p(x, y)$ represents the mean network gain—across mice and across sessions—within a 2D average window of width $[x - 3{:}x + 3]$° per s and height $[y - 15{:}y + 15]$°. A 2D Gaussian filter of s.d. = 15 (15° × 22.5° per s) is then applied. The network gain, the decoded HD and the measured HD were all smoothed with a moving-average filter of width 20 frames (~667 ms) while the measured head angular velocity was approximated by a simple linear regression with a regression window of similar width. To evaluate the significance of the difference between gain heat maps (Fig. 4f), we performed a Wilcoxon rank-sum test to compare, at each pixel, the gain distributions within the 2D window of width $[x - 3{:}x + 3]$° per s and height $[y - 15{:}y + 15]$° between darkness epochs of the 20 s experiment and D2 of the 2 min experiment. As we are only interested in the significance of the positive values (indicating the appearance of new bumps), negative values as well as $P > 0.001$ were marked as NaN.

Drift-speed heat-map analysis

Drift-speed heat maps were generated according to the same approach as for the gain heat maps. However, drift speed was approximated by a simple linear regression with a regression window of width 20 frames (~667 ms). The P-value matrix for drift speed difference between the 20 s experiment and D2 of the 2 min experiment (Extended Data Fig. 13e) was calculated as described above. However, only P values > 0.001 were marked as NaN.

Vector field analysis

The purpose of this analysis is to illustrate baseline attractiveness. We define the state space (y axis, drift-speed (° per s); x axis, drift-angle (°)). We construct a vector field matrix by dividing the x axis into 18 bins of width

20° each within the range [−180:180]°, and the y axis into 20 bins of width 0.03° per s each, within the range [−3:3]° per s. At each bin (x,y), we calculate the mean drift speed and mean drift acceleration across mice and across sessions. The two latter quantities represent the velocity components (u,v) that determine the length and direction of the velocity vector. We assume the vector field has a central symmetry with reference to the baseline point (0,0) owing to the symmetry in the experimental design. We therefore generate an image of the original vector field that is its reflection across the origin. The two versions are then averaged to produce the final 2D vector field. Streamlines are generated using the streamline function in MATLAB. For Figs. 4d and 5e, streamlines were simulated over 1,000 timesteps.

Analysis of cue-rotation sessions

At the beginning of each continuous cue-rotation epoch, the visual cue was displayed at the same location as in the baseline. After cue removal and depending on its previous rotation speed, the cue would have either reached ±180° or ±90° (cue orientation in clockwise-cue-rotation sessions was reflected across the x axis so that the cue ends at ±180° (fast cue-rotation) or −90° (slow cue-rotation)). The offset is therefore expected to start within a close range of these two directions during the second darkness epoch. However, in some cases, drifts during the first darkness epoch were large enough so that the initial anchoring to the rotating cue occurred considerably far from baseline. This caused the drift signal during the second darkness to start further away from the expected location. In Fig. 5d, we limited our analysis to drifts starting within [−180:−145]U[145:180]° for fast cue rotation and [−125:−55]° for slow cue rotation, to study the effects across sessions with similar stability during baseline (total $n = 44$ out of 60).

Dimensionality reduction

It is generally believed that the main function of the HD system is to provide an estimate of the HD at any given time. As most studies of this network, including ours, are conducted while recording the neural activity in animals placed on horizontal planes, it is fair to assume that most of the variability in the activity of HD neural population can be captured by a single variable representing the angle faced by the animal, at an instant t, with reference to a

given allocentric reference frame. Indeed, previous studies have shown that, in stable conditions, different dimensionality reduction methods[17,38] would produce a circular manifold that can be fairly approximated in a unidimensional polar state-space with a fixed radius. Nevertheless, a previous study[17] observed that the structure becomes more complex during slow-wave sleep. Our guiding hypothesis is that the intrinsic geometric structure of the neural activity in the HD network lies in a multidimensional state space and that latent variables other than the angular component are needed to explain the variability in spiking data, during non-stable conditions such as resets and drift situations. Here we propose the simplest augmentation to the latent structure by adding a radial component that we expect to indicate instantaneous changes in global firing activity of the HD network. Although we believe that the true intrinsic dimensionality of the HD neural data is higher than two, the current paper mainly focuses on the necessity of at least a second dimension of the HD system during instability.

To test our hypothesis, we developed a deep feedforward neural network that maps the high-dimensional input (neural) data onto the 2D polar space (angular dimension θ and radial dimension R). The network is trained on circular data from the measured head direction. The radial component R is a latent variable that can take any non-negative value. Our previous analyses (not included here) have shown that, although methods such as principal component analysis and Isomap can uncover looped latent structures, these unsupervised learning algorithms tend to produce distorted circles, in the presence of noise, when applied on baseline data (that is, stable condition), which makes the definition of a radius less straightforward and motivates our use of a supervised learning method.

We used a feedforward neural network with three parallel branches. Two of these branches have three fully connected hidden layers (referred to as first and second or B_1 and B_2, respectively), while the third branch has two fully connected hidden layers (referred to as middle or B_m) (Extended Data Fig. 4d). The input layer receives a N × 1 vector of neural activity from N ADN neurons at time t (both calcium traces as well as firing activity from deconvolved spikes can be fed to the model). The output layer is composed of two units that are the results of multiplying the output g_t of the middle branch with the output $z_{1,t}$ of the first branch, on one hand, and the output

$z_{2,t}$ of the second branch, on the other hand, as illustrated in the diagram of Extended Data Fig. 4d.

We trained our model on baseline data. The objective is to find the set of weights W that minimize the distance between the network output $\left(\begin{array}{c}{g}_{t}{z}_{1,t}\\ {g}_{t}{z}_{2,t}\end{array}\right)$ and the vector $\left(\begin{array}{c}\cos ({\theta }_{t})\\ \sin ({\theta }_{t})\end{array}\right)$, where θ_t is the measured head direction of the animal at instant t. We define the loss function as the mean squared error:

$${\rm{MSE}}=\frac{1}{T}\mathop{\sum }\limits_{t=1}^{T}{\parallel (\begin{array}{c}\cos ({\theta }_{t})\\ \sin ({\theta }_{t})\end{array})-(\begin{array}{c}{g}_{t}{z}_{1,t}\\ {g}_{t}{z}_{2,t}\end{array})\parallel }_{2}^{2},$$

where, T is the duration of the training epoch and $\|.\|_2$ is the L_2 norm. If the algorithm converges, we obtain the following approximations:

$$\left\{\begin{array}{c}{z}_{1,t}\approx \frac{\cos ({\theta }_{t})}{{g}_{t}}\\ {z}_{2,t}\approx \frac{\sin ({\theta }_{t})}{{g}_{t}}\end{array}\right..$$

Let ${\hat{R}}_{t}=\frac{1}{{g}_{t}}$, then we can rewrite the output of each branch:

$$\left\{\begin{array}{c}{B}_{1}:{z}_{1,t}\approx {\hat{R}}_{t}\cos \left({\theta }_{t}\right)\\ {B}_{2}:{z}_{2,t}\approx {\hat{R}}_{t}\sin \left({\theta }_{t}\right)\\ {B}_{{\rm{m}}}:{g}_{t}=\frac{1}{{\hat{R}}_{t}}\end{array}\right..$$

In effect, this would allow branches B_1 and B_2 to learn a mapping from the input (neural) space to the Cartesian transformation of the polar coordinates of a given state s_t, at any time t (respectively, B_1 projects the input onto the x axis and, B_2 projects the input onto the y axis). From these two branches, we can extract the decoded angle ${\hat{\theta }}_{t}=\arctan \left(\frac{{z}_{2,t}}{{z}_{1,t}}\right)$. While branch $B_{\rm m}$ would learn a

mapping from the input space to the inverse of the approximate radius ($\hat{R}_{t}$) of said state, in polar space. If we assume that ($\hat{R}_{t}$) is a certain reflection of global neural activity, as per our hypothesis, then we expect small fluctuations of population activity in the training data (baseline) to be sufficient to allow the network to extrapolate ($\hat{R}_{t}$) on test data with larger fluctuations.

Statistics and reproducibility

All statistical tests are noted where the corresponding results are reported throughout the main text and Supplementary Information. All tests were uncorrected two-tailed tests unless otherwise noted. Outliers were identified as data points that fall outside the mean ± (3 s.d.) range.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The complete dataset for all experiments is available at Figshare (https://doi.org/10.6084/m9.figshare.21792689). The dataset should not be used for republication without prior consent from the authors.

Code availability

All source codes used in the current study are available on request to the corresponding authors.

References

1. Taube, J. S. The head direction signal: origins and sensory-motor integration. *Annu. Rev. Neurosci.* **30**, 181–207 (2007).

2. Butler, W. N., Smith, K. S., van der Meer, M. A. A. & Taube, J. S. The head-direction signal plays a functional role as a neural compass during navigation. *Curr. Biol.* **27**, 1259–1267 (2017).

3. Skaggs, W. E., Knierim, J. J., Kudrimoti, H. S. & McNaughton, B. L. A model of the neural basis of the rat's sense of direction. *Adv. Neural. Inf. Process. Syst.* **7**, 173–180 (1995).

4. Taube, J. S., Muller, R. U. & Ranck, J. B. Jr. Head-direction cells recorded from the postsubiculum in freely moving rats. I. Description and quantitative analysis. *J. Neurosci.* **10**, 420–435 (1990).

5. Taube, J. S., Muller, R. U. & Ranck, J. B. Jr. Head-direction cells recorded from the postsubiculum in freely moving rats. II. Effects of environmental manipulations. *J. Neurosci.* **10**, 436–447 (1990).

6. Zugaro, M. B., Arleo, A., Berthoz, A. & Wiener, S. I. Rapid spatial reorientation and head direction cells. *J. Neurosci.* **23**, 3478–3482 (2003).

7. Zugaro, M. B., Berthoz, A. & Wiener, S. I. Background, but not foreground, spatial cues are taken as references for head direction responses by rat anterodorsal thalamus neurons. *J. Neurosci.* **21**, RC154 (2001).

8. Goodridge, J. P., Dudchenko, P. A., Worboys, K. A., Golob, E. J. & Taube, J. S. Cue control and head direction cells. *Behav. Neurosci.* **112**, 749–761 (1998).

9. Taube, J. S. & Bassett, J. P. Persistent neural activity in head direction cells. *Cereb. Cortex* **13**, 1162–1172 (2003).

10. Valerio, S. & Taube, J. S. Path integration: how the head direction signal maintains and corrects spatial orientation. *Nat. Neurosci.* **15**, 1445–1453 (2012).

11. Cullen, K. E. & Taube, J. S. Our sense of direction: progress, controversies and challenges. *Nat. Neurosci.* **20**, 1465–1473 (2017).

12. Dumont, J. R. & Taube, J. S. The neural correlates of navigation beyond the hippocampus. *Prog. Brain Res.* **219**, 83–102 (2015).

13. McNaughton, B. L., Chen, L. L. & Markus, E. J. "Dead reckoning," landmark learning, and the sense of direction: a neurophysiological and computational hypothesis. *J. Cogn. Neurosci.* **3**, 190–202 (1991).

14. Taube, J. S. Head direction cells and the neurophysiological basis for a sense of direction. *Prog. Neurobiol.* **55**, 225–256 (1998).

15. Goodridge, J. P. & Taube, J. S. Preferential use of the landmark navigational system by head direction cells in rats. *Behav. Neurosci.* **109**, 49–61 (1995).

16. Knierim, J. J., Kudrimoti, H. S. & McNaughton, B. L. Interactions between idiothetic cues and external landmarks in the control of place cells and head direction cells. *J. Neurophysiol.* **80**, 425–446 (1998).

17. Chaudhuri, R., Gercek, B., Pandey, B., Peyrache, A. & Fiete, I. The intrinsic attractor manifold and population dynamics of a canonical cognitive circuit across waking and sleep. *Nat. Neurosci.* **22**, 1512–1520 (2019).

18. Viejo, G. & Peyrache, A. Precise coupling of the thalamic head-direction system to hippocampal ripples. *Nat. Commun.* **11**, 2524 (2020).

19. Arleo, A. et al. Optic flow stimuli update anterodorsal thalamus head direction neuronal activity in rats. *J. Neurosci.* **33**, 16790–16795 (2013).

20. Zugaro, M. B., Tabuchi, E. & Wiener, S. I. Influence of conflicting visual, inertial and substratal cues on head direction cell activity. *Exp. Brain Res.* **133**, 198–208 (2000).

21. Zugaro, M. B. et al. Rat anterodorsal thalamic head direction neurons depend upon dynamic visual signals to select anchoring landmark cues. *Eur. J. Neurosci.* **20**, 530–536 (2004).

22. Jayakumar, R. P. et al. Recalibration of path integration in hippocampal place cells. *Nature* **566**, 533–537 (2019).

23. Cope, A. J., Sabo, C., Vasilaki, E., Barron, A. B. & Marshall, J. A. A computational model of the integration of landmarks and motion in the insect central complex. *PLoS ONE* **12**, e0172325 (2017).

24. Page, H. J. I. & Jeffery, K. J. Landmark-based updating of the head direction system by retrosplenial cortex: a computational model. *Front. Cell Neurosci.* **12**, 191 (2018).

25. Ocko, S. A., Hardcastle, K., Giocomo, L. M. & Ganguli, S. Emergent elasticity in the neural code for space. *Proc. Natl Acad. Sci. USA* **115**, E11798–E11806 (2018).

26. Fisher, Y. E., Lu, J., D'Alessandro, I. & Wilson, R. I. Sensorimotor experience remaps visual input to a heading-direction network. *Nature* **576**, 121–125 (2019).

27. Kim, S. S., Hermundstad, A. M., Romani, S., Abbott, L. F. & Jayaraman, V. Generation of stable heading representations in diverse visual scenes. *Nature* **576**, 126–131 (2019).

28. Aharoni, D. & Hoogland, T. M. Circuit investigations with open-source miniaturized microscopes: past, present and future. *Front. Cell Neurosci.* **13**, 141 (2019).

29. Cai, D. J. et al. A shared neural ensemble links distinct contextual memories encoded close in time. *Nature* **534**, 115–118 (2016).

30. Ghosh, K. K. et al. Miniaturized integration of a fluorescence microscope. *Nat. Methods* **8**, 871–878 (2011).

31. Friedrich, J., Zhou, P. & Paninski, L. Fast online deconvolution of calcium imaging data. *PLoS Comput. Biol.* **13**, e1005423 (2017).

32. Zhou, P. et al. Efficient and accurate extraction of in vivo calcium signals from microendoscopic video data. *eLife* **7**, e28728 (2018).

33. Peyrache, A., Lacroix, M. M., Petersen, P. C. & Buzsaki, G. Internally organized mechanisms of the head direction sense. *Nat. Neurosci.* **18**, 569–575 (2015).

34. Taube, J. S. Head direction cells recorded in the anterior thalamic nuclei of freely moving rats. *J. Neurosci.* **15**, 70–86 (1995).

35. Blair, H. T., Lipscomb, B. W. & Sharp, P. E. Anticipatory time intervals of head-direction cells in the anterior thalamus of the rat: implications for path integration in the head-direction circuit. *J. Neurophysiol.* **78**, 145–159 (1997).

36. Seelig, J. D. & Jayaraman, V. Neural dynamics for landmark orientation and angular path integration. *Nature* **521**, 186–191 (2015).

37. Wei, X. X. et al. A zero-inflated gamma model for post-deconvolved calcium imaging traces. *Neurons, Behavior, Data Analysis, and Theory* https://doi.org/10.48550/arXiv.2006.03737 (2020).

38. Rubin, A. et al. Revealing neural correlates of behavior without behavioral measurements. *Nat. Commun.* **10**, 4745 (2019).

39. Zugaro, M. B., Tabuchi, E., Fouquier, C., Berthoz, A. & Wiener, S. I. Active locomotion increases peak firing rates of anterodorsal thalamic head direction cells. *J. Neurophysiol.* **86**, 692–702 (2001).

40. Redish, A. D. *Beyond the Cognitive Map: From Place Cells to Episodic Memory* (MIT Press, 1999).

41. Hulse, B. K. & Jayaraman, V. Mechanisms underlying the neural computation of head direction. *Annu. Rev. Neurosci.* **43**, 31–54 (2020).

42. Kim, S. S., Rouault, H., Druckmann, S. & Jayaraman, V. Ring attractor dynamics in the *Drosophila* central brain. *Science* **356**, 849–853 (2017).

43. Tolman, E. C. Cognitive maps in rats and men. *Psychol. Rev.* **55**, 189–208 (1948).

44. O'Keefe, J. & Nadel, L. *The Hippocampus as a Cognitive Map* (Clarendon Press, 1978).

45. Pnevmatikakis, E. A. & Giovannucci, A. NoRMCorre: an online algorithm for piecewise rigid motion correction of calcium imaging data. *J. Neurosci. Methods* **291**, 83–94 (2017).

46. Pnevmatikakis, E. A. et al. Simultaneous denoising, deconvolution, and demixing of calcium imaging data. *Neuron* **89**, 285–299 (2016).

47. Zhang, K., Ginzburg, I., McNaughton, B. L. & Sejnowski, T. J. Interpreting neuronal population activity by reconstruction: unified framework with application to hippocampal place cells. *J. Neurophysiol.* **79**, 1017–1044 (1998).

48. Jackson, J. C. & Redish, A. D. Detecting dynamical changes within a simulated neural ensemble using a measure of representational quality. *Network* **14**, 629–645 (2003).

Acknowledgements

We thank S. Kim, H. C. Yong and A. Nieto-Posadas for the technical assistance and help gathering the histology data; L. Paninski, M. Hasselmo, S. Badrinarayanan, J. Ying, M. Yaghoubi and A. Peyrache for comments on previous versions of this paper; D. Aharoni for advice and assistance with the UCLA miniscope; and the members of the Brandon laboratory for discussions, in particular, M. Yaghoubi and R. R. Rozeske. This work was supported by funding from the Canadian Institutes of Health Research (project grants 367017 and 377074), the Natural Sciences and Engineering and Research Council of Canada (Discovery 74105) and the Canada Research Chairs Program to M.P.B.

Author information

Author notes

1. Zaki Ajabi

Present address: Department of Neurobiology, Harvard Medical School, Boston, MA, USA

Authors and Affiliations

1. Department of Psychiatry, Douglas Hospital Research Centre, McGill University, Verdun, Quebec, Canada

   Zaki Ajabi, Alexandra T. Keinath & Mark P. Brandon

2. Integrated Program in Neuroscience, McGill University, Montreal, Quebec, Canada

   Zaki Ajabi & Mark P. Brandon

3. Department of Neuroscience, University of Texas at Austin, Austin, TX, USA

   Xue-Xin Wei

4. Department of Psychology, University of Texas at Austin, Austin, TX, USA

   Xue-Xin Wei

5. Center for Perceptual Systems, University of Texas at Austin, Austin, TX, USA

   Xue-Xin Wei

6. Center for Theoretical and Computational Neuroscience, University of Texas at Austin, Austin, TX, USA

   Xue-Xin Wei

Contributions

Z.A., A.T.K. and M.P.B. contributed to the experimental design. Z.A. performed all surgeries, recordings, data analysis and modelling. X.-X.W.,

A.T.K. and M.P.B. provided guidance for data analysis. Z.A. wrote the initial draft. All of the authors contributed to editing and revising the paper.

Corresponding authors

Correspondence to Zaki Ajabi or Mark P. Brandon.

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks Kate Jeffery and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Calcium imaging in the anterodorsal thalamic nucleus (ADN) and identification of HD neurons.

a. Histology data showing coronal brain sections from each mouse with GCaMP6f expression, in ADN (anterior part). Mouse ID written in the top right and scale-bars shown in the bottom left of each panel. In total, 12 mice were injected and implanted for this study, only 3 (shown here) provided

enough simultaneously recorded head-direction cells for continued experimentation. b. Directional maps of ADN in each mouse. HD cells are coloured by their preferred firing direction (PFD). Colour-wheel shows angle-colour assignments. Mouse ID written on the top right and scale-bars shown in the bottom left of each panel. c. Examples of HD cells' coverage of the azimuthal plane, in each mouse. Rows in each matrix represent tuning curve heatmaps of individual HD cells. The amplitudes of individual tuning curves are normalized. Mouse ID written above each panel. d. Left: An example polar tuning curve for a HD neuron. Yellow line: direction of maximum firing activity (that is, PFD). Firing activity is occupancy normalized. Right: Top-row: Example calcium signal deltaF/F (green) from one HD neuron and deconvolved trace (red). Both traces were normalized. Middle-row: Measured HD. Bottom-row: Extracted stimulus signal of the HD neuron's PFD. Peaks indicate instances of the animal facing the particular PFD. The deconvolved signal is cross-correlated with the stimulus signal in order to obtain the Pearson's correlation coefficient which reflects the cell's degree of HD tuning ($r = 0.85$ in the case of the current example). e. Distributions of correlation coefficients after 1000 circular-shift shuffles of the firing activity signals (smoothed deconvolved traces) of all HD neurons, in each mouse. Red and green vertical lines indicate 95th and 99th percentiles, respectively. Data includes 10 baseline recordings of 3 min each, for every mouse. Of all recorded cells, ~94% met the 95th percentile selection criterion while ~83% met the 99th percentile selection criterion.

Extended Data Fig. 2 Polar tuning curves of ADN neurons from a 10-minute baseline recording for each mouse (total number of neurons = 502).

The directional tuning of each ADN neuron is shown by the correlation coefficients above each tuning curve. Mouse ID written on the left side.

Extended Data Fig. 3 Anticipatory behaviour and drift-speed pattern during baseline.

a. Top row: Mean bump of activity divided between positive (blue) and negative (pink) head angular velocities (HAV). Bar graph: Mean difference

between measured and decoded HD (n = 42x5000 data points from baseline recordings, between both groups; Two-sided Wilcoxon signed rank test: HAV < 0: p = 0, Z = 83.71; HAV > 0: p = 0, Z = −76.81). Bottom row: Mean cross-correlation of the mean bump of activity, per epoch, with the mean bump of activity for positive (blue) and negative (pink) HAVs. Bar graph: Mean peak angle of cross-correlation (n = 42x5000 data points from baseline recordings, between both groups; Two-sided Wilcoxon signed rank test: HAV < 0: p = 0, Z = −115.24; HAV > 0: p = 0, Z = 113.13). Both analyses show a significant amount of anticipation of future heading by the HD network. b. Top: Drift-speed heatmap showing an increased latency in updating the internal representation as the HAV becomes larger. Bottom: same pattern as the above, seen here in Internal HAV-versus-Measured HAV space. Notice the deviations of the mean signal (orange) from the diagonal, at high measured HAVs. Bar graphs indicate mean ± SEM.

Extended Data Fig. 4 Hypothesized relationship between the population-activity and movements in the low dimensional polar state-space.

a, **b**. The amount of the change of neural activity during bump movement depends on the gain of the network. *x* axis represents the neuronal space (assuming uniform distribution of HD cells by PFD). Mathematically, the distance between representations of internal HD from start to end of a rotation, in the Euclidean sense, is smaller at lower network gain. *D*: Euclidean distance, ${\boldsymbol{r}}_{t}^{activity}$: Nx1 vector of firing rates from N HD neurons, at time *t* for ‘high’ or ‘low’ activity levels. **c**. The concept of decreasing distance between internal HD representations, at lower network gain, is naturally captured in the 2D polar plane if we assume that radius reflects the level of network activity. The distance travelled in the hypothetical state-space of the HD network is greater when the radius is larger as well as when the net gain is higher, which could be quantified by the total change of firing rate across the network. Thus, we hypothesize that radius is correlated with overall population activity (that is, network gain) and that decreasing distance facilitates rotations across the HD network. Assuming that the internal HD representation lives in a 2D polar state-space where each state is defined by phase and radius, state transitions would be fastest at the lower end of the radial component because

of the decreasing distance between states representing different angles, near the centre of the baseline ring. Bar-graphs are only indicative and not to scale. **d**. Diagram of the artificial neural network used to project high-dimensional neural activity onto 2D polar space. Numbers inside each box correspond to the unit count. All activation functions are ‘relu’ except for nodes $z_{1,t}$ and $z_{2,t}$ where the activation function is ‘tanh’. In all layers, we apply L_2 regularization with regularization factor 0.001. Input data is normalized.

Extended Data Fig. 5 Relationship between network gain and population activity.

a. Reconstructed bump of activity (averaged over n = 42 sessions of the first experiment) for varying network gain ranges. Gain modulation not only affects the activity packet but also baseline activity. The decreasing baseline amplitude at low network gain indicates that the modulation is not driven by increased activity outside the main activity packet. Notice that the width of the activity packet remains within a narrow range. ‘fwhm’: the full width at half maximum in °. b. Method used to determine the variance explained by gain. Using the internal HD and neural activity from all recorded neurons per session as inputs ($S_{neuron\ i}$; 5 examples shown for illustration purposes), we can extract the tuning curve of each neuron (average firing activity as a function of internal HD, $f(\theta_t)$) as well as the gain signal (g_t), while assuming that pairwise coherence between HD cells is preserved. Two reconstructions of the neural activity are then produced from tuning curves and internal HD: In the first case (Dark-blue) neural activity is multiplied by gain ($R_{neuron\;i}^{g}$) while in the second case (Light-blue), gain is not taken into account ($R_{neuron\ i}$). The sum of variance across neurons is calculated for each group of neural activity (including ground-truth ($S_{neuron\ i}$)). c. Comparison of variance explained in percentage between the neural activity reconstruction with and without gain (sum of variance in each group is divided by the sum of variance in the ground-truth group) (n = 42 sessions; Two-sided Wilcoxon rank-sum test: p = 0.0245, Z = 2.2499). Error bars show mean ± SEM. d. Increase in variance explained when gain is applied to reconstructed neural activity relative to the case where gain is not applied (that is, ratio between % variance explained with and without gain,

minus 1) (n = 42 sessions; Mean = 13.71%, s.d. = 5.14%). Error-bars show mean ± s.d. Dots represent individual datapoints.

Extended Data Fig. 6 Reset behaviour and gain modulation across mice.

a. Top row: Resets separated according to their speeds between fast (light blue) and slow (dark blue) groups. Mouse ID written above each panel. Bottom row: Corresponding gain signals for fast and slow resets. b. Top row: Resets separated according to their range between long- (light blue), mid- (dark blue) and short- (grey) groups. Mouse ID written above each panel. Bottom row: Corresponding gain signals for long-, mid- and short-range resets. Data are mean ± s.e.m.

Extended Data Fig. 7 Agreement between true and model-predicted resets.

a. Averaged heatmaps of the reconstructed bump of activity during fast (left column) and slow (right column) resets (same data as in Fig. 2g,h). Data is presented in the egocentric reference frame, without drift adjustment (top row) and with drift adjustment (bottom row) showing, in both cases, no additional bumps outside the main activity packet. Dashed red line indicates cue-onset, while white horizontal line at 90° is for reference. Firing activity is normalized. b. Simulation output of the gain-modulated attractor model taking input data (that is, gain) as in a. c. Top: Mean simulated reset signals for fast (light blue) and slow (dark blue) groups. Bottom: Mean simulated gain signals for the same groups. Data are mean ± s.e.m. Dashed signals represent means of ground-truth data. d. Individual examples of simulation predictions (red lines) for fast and slow reset groups, plotted against actual resets (blue lines). Yellow lines indicate cue location. Amplitudes are relative to angles at cue-onset (dashed black line).

Extended Data Fig. 8 Animal behaviour, prior to cue display, is predictive of reset speed.

a. Triggered average of gain shows a sharp decrease after cue display (Two-sided Wilcoxon rank-sum test: average gain 1-second pre-cue versus average

gain 1-second post-cue: $p = 0.0228$, $Z = 2.28$) (top). However, overall absolute head angular velocity (aHAV) does not seem to differ before and after cue display (Two-sided Wilcoxon rank-sum test: average aHAV 1-second pre-cue versus average aHAV 1-second post-cue: $p = 0.6259$, $Z = 0.49$) (bottom). Same reset events as in Fig. 2g,h (n = 42 events). b. Separation of signals in a. between fast (Light blue; n = 22 events) and slow (Dark blue; n = 20 events) resets shows similar gain amplitudes over a 1-second interval prior to cue display (Two-sided Wilcoxon rank-sum test: $p = 0.3580$, $Z = 0.92$) (top). However, aHAV is lower for fast resets compared with slow resets, over the same period (Two-sided Wilcoxon rank-sum test: $p = 0.0294$, $Z = 2.18$) (Bottom). c. Head angular velocity becomes more predictive of reset type closer to the moment of cue-display when compared with prediction performance based on gain amplitudes within the same time interval. Deviance of the fit is used as defined in Matlab's mnrfit function for logistic regression. Data shown is same as in Fig. 2g,h. Time dependent signals, in a and b, are shown as mean ± s.e.m. and bar-graphs show mean ± s.e.m. with individual datapoints.

Extended Data Fig. 9 Relationship between reset range and gain modulation.

a. Mean drifts for short- (grey; n = 27 events), mid- (dark blue; n = 40 events) and long- (light blue; n = 67 events) range reset-groups showing non-significant difference in drift-speeds between mid- and long-range groups (Two-sided Wilcoxon rank-sum test: Short-Mid: $p = 4.19e-5$, $Z = 4.10$; Short-Long: $p = 7.73e-5$, $Z = 3.95$; Mid-Long: $p = 0.62$, $Z = 0.50$; 150 frames (~5 s) post-cue). b. Network gains for the short-, mid- and long-ranges have similar amplitudes prior to cue-display (Two-sided Wilcoxon rank-sum test: Short-Mid: $p = 0.1174$, $Z = 1.57$; Short-Long: $p = 0.32$, $Z = 1.00$; Mid-Long: $p = 0.2984$, $Z = 1.04$; 50 frames (~1.67 s) pre-cue), yet they exhibit gradual decrease after cue-display (Two-sided Wilcoxon rank-sum test: Short-Mid: $p = 0.0129$, $Z = 2.49$; Short-Long: $p = 2.6876e-9$, $Z = 5.95$; Mid-Long: $p = 1.2130e-5$, $Z = 4.38$; 150 frames (~5 s) post-cue). c. Relationship between average gain and reset range. Each dot represents a correct reset event (n = 134 events). The R^2 value corresponds to a linear regression model fit (green line). All clockwise sessions have been reflected across the x-axis and transformed into counter-clockwise ones. d. Rapid gain

spikes can be seen shortly after cue-display, in the three reset-range groups (Same data as in b, with higher temporal resolution). All reset ranges start at similar amplitudes at the end of the darkness period (Two-sided Wilcoxon rank-sum test: short-mid: $p = 0.3940$, $Z = 0.85$; short-long: $p = 0.2090$, $Z = 1.26$; mid-long: $p = 0.4686$, $Z = 0.72$). Following cue-display, each group exhibits a brief gain increase (5 frames (~150 ms) pre-cue vs 5 frames (~150 ms) post-cue: Two-sided Wilcoxon rank-sum test: short: $p = 6.9690e-4$, $Z = 3.39$; mid: $p = 0.0369$, $Z = 2.09$; long: $p = 2.6898e-4$, $Z = 3.64$). These gain spikes are largest for the short-range group (Two-sided Wilcoxon rank-sum test: short-mid: $p = 4.4888e-4$, $Z = 3.51$; short-long: $p = 1.8600e-4$, $Z = 3.74$; mid-long: $p = 0.9326$, $Z = 0.08$). Time-dependent signals are shown as data are mean ± s.e.m. and bar-graphs show mean ± s.e.m. with individual datapoints.

Extended Data Fig. 10 Distinct drift and gain patterns across darkness periods.

a. Drift variability increases significantly following a reset (D2, D3 and D4) in comparison with D1 (Mean drift s.d. compared across darkness epochs: Two-sided Wilcoxon rank-sum test: BL-D1: $p = 3.1214e-15$, $Z = 7.89$; D1-D2: $p = 1.1477e-6$, $Z = 4.86$; D1-D3: $p = 8.3761e-5$, $Z = 3.93$; D1-D4: $p = 5.6600e-11$, $Z = 6.55$). Drift s.d. also increases with time after a reset (D2, D3 and D4) while it remains constant following baseline (D1). (Number of epochs: D1: $n = 42$; D2: $n = 35$; D3: $n = 32$; D4: $n = 35$). b. Mean drift-speed in each darkness epoch shows systematic biases that depend on prior cue-event. (Two-sided Wilcoxon rank-sum test: BL-D1: $p = 0.1250$, $Z = 1.53$; Two-sided Wilcoxon signed rank test: D2: $p = 0.0168$, $Z = -2.39$; D3: $p = 0.0313$, $Z = 2.15$; D4: $p = 2.9929e-4$, $Z = 3.62$). (Number of epochs: D1: $n = 42$; D2: $n = 35$; D3: $n = 33$; D4: $n = 34$). c. Comparison between drifts in D2 and D4 of the 90°-cue-shift experiment. Although the two events are experimentally symmetric to each other with reference to baseline, drifts in D4 appear to have larger biases (in absolute value terms) than D2. Left: Mean drift signals, in D2 (green) and D4 (dark-blue). Drifts in D2 have been mirrored across the 0°-line for comparison purposes. Right: Comparison between average drift speeds, in D2-mirrored (green; $n = 35$ epochs) and D4 (dark-blue; $n = 34$ epochs) (Two-sided Wilcoxon rank-sum test: $p = 0.0184$, $Z = 2.36$). d. Average gain tuning curves across light conditions. e. Average

gain tuning curves across darkness conditions show a gradual decrease of the network gain away from the internal cue location (dashed yellow line) from D1 to D4 (Number of epochs: D1: n = 42; D2: n = 35; D3: n = 33; D4: n = 35). Time-dependent signals and gain tuning curves are shown as mean ± s.e.m. bar-graphs show mean ± s.e.m. with individual datapoints.

Extended Data Fig. 11 Network gain patterns across mice and darkness epochs.

a. Network gain during darkness shown as heatmaps (top row) and tuning curves (bottom row), per mouse. In both cases, data is averaged across sessions and darkness epochs (D1 to D4) of the 90°-cue-shift experiment. Values for the tuning curves are shown as mean ± s.e.m. Mouse ID written above each panel. b. Top row: Network gain heatmaps showing same data as in a, split (from left to right, respectively) across the different darkness epochs D1 to D4 of the 90°-cue-shift experiment. Bottom row: Drift speed heatmaps showing a consistent pattern, yet with varying amplitudes, across darkness epochs D1 to D4. No obvious effect of the gain landscape can be seen in these patterns and gain fluctuations did not correlate with any measurable distortion to the drift-speed landscape within the Head AV-vs-Internal HD state-space which maintained similar patterns to baseline (Extended Data Fig. 3b). This observation draws a clear distinction from the rapid representational shifts seen during resets and may point to a completely different mechanism linking network gain and drifts in dark conditions.

Extended Data Fig. 12 Model of the hypothesized role of plasticity in baseline attraction during darkness.

a. To explain drift dynamics that we observed in darkness, we propose a model that incorporates a ‘sensorimotor-by-HD’ layer that represents a cortical consensus about the directional sensory experience. Each neuron in this layer synapses onto all HD neurons via plastic synapses. Depending on the duration of exposure to the shifted cue context, the network either has enough time (that is, 2 min case) to form new associations between neurons of the HD and sensorimotor-by-HD layers which results in the emergence of a new steady state and no reversion to baseline, or not enough time (that is,

20 s case) and so, baseline associations between the two layers are maintained which causes the internal HD representation to revert to baseline state. b, c, d, and e. Model simulations. b. Synaptic weight matrix linking the HD layer to the Sensorimotor-by-HD layer, during baseline. c. Simulations of representational drifts in 20 s (bottom row), 2 min (middle row) and 4 min (top row) exposures to the reset context. Behaviour for individual examples (that is, head angular velocity) is shared across scenarios and is taken from actual recordings. d. Synaptic weight matrices in darkness for the three scenarios showing the strengthening of new associations between HD and Sensorimotor-by-HD layers while baseline connections become weaker with increased duration of exposure to the reset context. e. Mean drifts (solid lines) in darkness across scenarios shaded areas indicate SEM.

Extended Data Fig. 13 Drift and gain patterns during reversion.

a. Mean drift signal during reversion ($n = 43$ epochs). Dashed yellow line divides the darkness period in two halves with contrasting states of the HD network: drifting (1st half) and stabilizing (2nd half). Data shown as mean ± s.e.m. Bar graph: Comparison of mean drift-speeds between the first and second halves of the darkness period (Two-sided Wilcoxon rank-sum test: $p = 3{,}5802e\text{-}8$, $Z = 5.51$). Data are mean ± s.e.m. with individual datapoints. b. Top row: Heatmaps of network gain during the first (left) and second (right) halves. Bottom row: Heatmaps of drift-speed during the first (left) and second (right) halves showing state-dependent distortions of the drift-speed pattern. No obvious relationship between drift speed and network gain landscapes could be determined, unlike what we observed during reset events, indicating that the relationship between gain and network state updating depends on the particular external input and/or current regime of the network. c, d, and e. Comparison of drift patterns between darkness epochs of the 20 s cue-exposure experiment and D2 of the 2 min cue-exposure experiment. c. Same as main Fig. 4c. d. Drift-speed heatmaps. Left: 20 s cue-exposure experiment ($n = 43$ epochs). Right: D2 of the 2 min cue-exposure experiment ($n = 35$ epochs). e. Left: Drift-speed difference (same data as in d) showing a significant distortion of the pattern seen in the first experiment around the internal location of the cue. Right: p-value matrix for data in left (Wilcoxon rank-sum test; pixels where $p > 0.001$ were marked as NaN). In addition to the network gain, the drift pattern also

shows systematic differences as a function of angular velocity and internal head direction between D2 and the darkness following 20 s visual-cue display.

Extended Data Fig. 14 Persistent drift biases in darkness after cue rotation, between actual data and model-based simulations.

a. Recorded examples of drift biases for continuous fast (left group) and slow (right group) cue-rotation. b, c, d and e. Model-simulation of vestibular input recalibration by visual experience. b. Synaptic weight matrix linking the HD layer to the Sensorimotor-by-HD layer (see model in Extended Data Fig. 12), during baseline. c. Simulations of offset during cue-rotation and in subsequent darkness for the fast (3°/s) and slow (1.5°/s; 1.28°/s) cases. Behaviour for individual examples (i.e. head angular velocity) is shared across scenarios and is taken from actual recordings. Sessions without vestibular input recalibration (that is, vestibular angular velocity neurons do not receive input from the bias cells – see model details in Supplementary Information) for both 3°/s and 1.5°/s cases were used as test examples. The 1.28°/s cue-rotation sessions were used to show the effect of cue rotation-speed on drift biases regardless of offset proximity to baseline condition. d. Synaptic weight matrices at the beginning of the 2nd darkness phase for fast (3°/s) and slow (1.5°/s) scenarios showing that baseline associations remain dominant even after 7 min of cue rotation which explains the stabilization around the 0°-offset line. e. Mean drifts (solid lines) in darkness across scenarios. Shaded areas indicate s.e.m.

Supplementary information

Supplementary Information

Supplementary Figs. 1–6 and Supplementary References.

Reporting Summary

Peer Review File

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Ajabi, Z., Keinath, A.T., Wei, XX. *et al.* Population dynamics of head-direction neurons during drift and reorientation. *Nature* **615**, 892–899 (2023). https://doi.org/10.1038/s41586-023-05813-2

- Received: 16 September 2021
- Accepted: 03 February 2023
- Published: 22 March 2023
- Issue Date: 30 March 2023
- DOI: https://doi.org/10.1038/s41586-023-05813-2

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

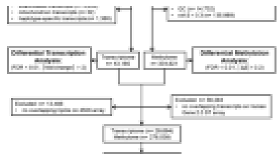
Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05813-2

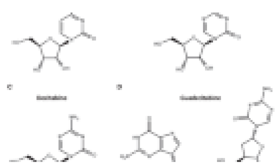
| Section menu | Main menu |



Combination treatment of acute myeloid leukemia cells with DNMT and HDAC inhibitors: predominant synergistic gene downregulation associated with gene body demethylation

23 November 2018

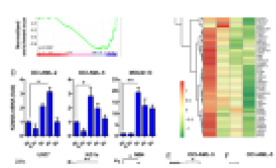
Nadja Blagitko-Dorfs, Pascal Schlosser, … Michael Lübbert



Hypomethylating agents (HMA) for the treatment of acute myeloid leukemia and myelodysplastic syndromes: mechanisms of resistance and novel HMA-based therapies

06 May 2021

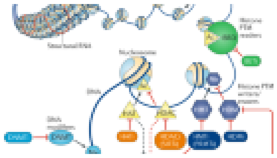
Julia Stomper, John Charles Rotondo, … Michael Lübbert



KDM6 demethylases integrate DNA repair gene regulation and loss of KDM6A sensitizes human acute myeloid leukemia to PARP and BCL2 inhibition

31 January 2023

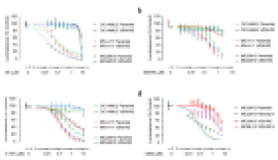
Liberalis Debraj Boila, Subhadeep Ghosh, … Amitava Sengupta



Combining epigenetic drugs with other therapies for solid tumours — past lessons and future promise

30 September 2019

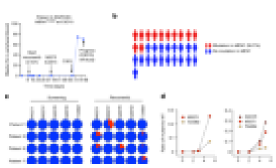
Daphné Morel, Daniel Jeffery, … Sophie Postel-Vinay



Activation of RAS/MAPK pathway confers MCL-1 mediated acquired resistance to BCL-2 inhibitor venetoclax in acute myeloid leukemia

21 February 2022

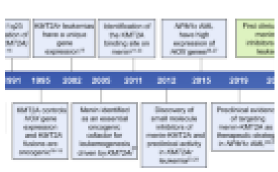
Qi Zhang, Bridget Riley-Gillis, … Marina Konopleva



MEN1 mutations mediate clinical resistance to menin inhibition

15 March 2023

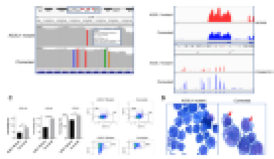
Florian Perner, Eytan M. Stein, … Sheng F. Cai



[Therapeutic implications of menin inhibition in acute leukemias](#)

15 June 2021

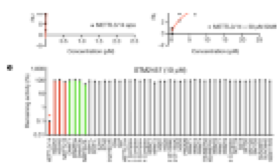
Ghayas C. Issa, Farhad Ravandi, … Michael Andreeff



[ASXL1 mutations are associated with distinct epigenomic alterations that lead to sensitivity to venetoclax and azacytidine](#)

21 September 2021

Nora E. Rahmani, Nandini Ramachandra, … Amit Verma



[Small-molecule inhibition of METTL3 as a strategy against myeloid leukaemia](#)

26 April 2021

Eliza Yankova, Wesley Blackaby, … Tony Kouzarides

- Article
- Open Access
- Published: 15 March 2023

The menin inhibitor revumenib in *KMT2A*-rearranged or *NPM1*-

mutant leukaemia

- Ghayas C. Issa ,
- Ibrahim Aldoss,
- John DiPersio,
- Branko Cuglievan,
- Richard Stone,
- Martha Arellano,
- Michael J. Thirman,
- Manish R. Patel,
- David S. Dickens,
- Shalini Shenoy,
- Neerav Shukla,
- Hagop Kantarjian,
- Scott A. Armstrong,
- Florian Perner,
- Jennifer A. Perry,
- Galit Rosen,
- Rebecca G. Bagley,
- Michael L. Meyers,
- Peter Ordentlich,
- Yu Gu,
- Vinit Kumar,
- Steven Smith,
- Gerard M. McGeehan &
- …
- Eytan M. Stein

Show authors

Nature **volume 615**, pages 920–924 (2023)

- 52k Accesses

- 1226 Altmetric

- Metrics details

Abstract

Targeting critical epigenetic regulators reverses aberrant transcription in cancer, thereby restoring normal tissue function[1,2,3]. The interaction of menin with lysine methyltransferase 2A (KMT2A), an epigenetic regulator, is a dependence in acute leukaemia caused by either rearrangement of *KMT2A* or mutation of the nucleophosmin 1 gene (*NPM1*)[4,5,6]. *KMT2A* rearrangements occur in up to 10% of acute leukaemias and have an adverse prognosis, whereas *NPM1* mutations occur in up to 30%, forming the most common genetic alteration in acute myeloid leukaemia[7,8]. Here, we describe the results of the first-in-human phase 1 clinical trial investigating revumenib (SNDX-5613), a potent and selective oral inhibitor of the menin–KMT2A interaction, in patients with relapsed or refractory acute leukaemia (ClinicalTrials.gov, NCT04065399). We show that therapy with revumenib was associated with a low frequency of grade 3 or higher treatment-related adverse events and a 30% rate of complete remission or complete remission with partial haematologic recovery (CR/CRh) in the efficacy analysis population. Asymptomatic prolongation of the QT interval on electrocardiography was identified as the only dose-limiting toxicity. Remissions occurred in leukaemias refractory to multiple previous lines of therapy. We demonstrate clearance of residual disease using sensitive clinical assays and identify hallmarks of differentiation into normal haematopoietic cells, including differentiation syndrome. These data establish menin inhibition as a therapeutic strategy for susceptible acute leukaemia subtypes.

Main

The prognosis of acute leukaemias harbouring rearrangements of the gene lysine methyltransferase 2A (*KMT2A*), previously known as mixed-lineage leukaemia (*MLL*), is poor, with a 5-year overall survival of less than 25% (ref. [7]). *KMT2A* rearrangements (*KMT2A*r) occur in 80% of infant acute lymphoblastic leukaemia (ALL) and in 5–15% of children and adults with acute leukaemia, whether myeloid, lymphoid or mixed phenotype[9]. Mutated nucleophosmin 1 gene (*NPM1*) is the most common genetic alteration in adult acute myeloid leukaemia (AML), occurring in up to 30% of patients[8].

Currently there are no targeted therapies specifically approved for acute leukaemia with *KMT2A*r or mutated *NPM1*.

Genetic rearrangements of *KMT2A* lead to aberrant expression of homeobox (*HOX*) genes and their DNA-binding cofactor *MEIS1* (ref. [10]). This gene expression programme, normally expressed in stem cells, causes a haematopoietic differentiation block and leukaemic transformation[11]. For leukaemias driven by *KMT2A*r, menin is a critical oncogenic cofactor[4]. The menin-binding motif is preserved throughout all KMT2A fusion proteins, and menin is an essential cofactor for binding to *HOX* gene promoters[7]. Similarly, in AML with mutated *NPM1*, the interaction between wild-type *KMT2A* and menin leads to *HOX*- and *MEIS1*-mediated leukaemogenic transcription[6,12]. Blockade of the menin–KMT2A interaction disrupts the assembly of oncogenic KMT2A wild-type or fusion complexes on chromatin[12]. Preclinical studies showed that menin inhibition downregulates *HOX* and *MEIS1* transcription and reverses leukaemogenesis in *KMT2Ar-r-* or *NPM1*-mutated leukaemia models[5,12,13]. Revumenib, previously known as SNDX-5613, is a potent, oral, selective inhibitor of the menin–KMT2A interaction. Treatment with revumenib led to abrogation of aberrant *HOX* gene expression and dramatic antileukaemic activity in those models (Extended Data Fig. 1; GEO accession no. for RNA sequencing (RNA-seq) data, GSE216730). In this first-in-human clinical trial we assessed the safety, maximum tolerated dose (MTD), recommended phase 2 dose (RP2D) and pharmacokinetic and pharmacodynamic profiles in patients with relapsed or refractory acute leukaemia, and present the clinical activity of revumenib in patients with *KMT2A*r or mutated *NPM1*.

This phase 1 dose-escalation study was conducted across nine sites in the United States. Because revumenib is a substrate of cytochrome P450 3A4 (CYP3A4), two parallel dose-escalation cohorts, one without (Arm A) and one with (Arm B) strong CYP3A4 inhibitors, were conducted. Revumenib was administered orally every 12 h (q12h) in continuous 28-day cycles. Although all patients could enrol at study initiation regardless of cytogenetic and mutational profile, given the strong preclinical rationale demonstrating that the menin–KMT2A interaction is a targetable vulnerability in acute leukaemia with *KMT2A*r or mutated *NPM1*, an early amendment to the protocol restricted enrolment to patients with relapsed or refractory acute

leukaemia with *KMT2A*r or mutated *NPM1* (88% of the total study population). Between 5 November 2019 and 31 March 2022, a total of 68 patients (37 in Arm A and 31 in Arm B) were enroled and treated. The data cutoff date for this analysis was 31 March 2022. Patient disposition is presented in Extended Data Fig. 2.

Baseline characteristics of patients are provided in Extended Data Table 1. There were 56 patients (82%) with relapsed or refractory AML, 11 (16%) with ALL and one with mixed-phenotype acute leukaemia (2%). Forty-six patients (68%) had *KMT2A*r, 14 (21%) had mutated *NPM1* and eight (12%) had neither *KMT2A*r nor *NPM1* mutations. Sixty patients were adults (at least 18 years old) and eight were children or adolescents (under 18 years of age). The median age was 42.5 years (range, 0.8–79). The median age of adult patients was 50.5 years (range, 19–79) and the median age among paediatric patients was 2.5 years (range, 0.8–16). The study population was heavily pretreated with a median of four previous lines of therapy (range, 1–12), and 31 patients (46%) had relapsed after an allogenic stem cell transplant. All patients who received at least one dose of revumenib were included in the safety analysis whereas only patients with *KMT2A*r or mutated *NPM1* were included in the efficacy analysis.

The dose-escalation schema is shown in Extended Data Fig. 3. The only dose-limiting toxicity observed in both Arm A and Arm B was grade 3 prolongation of the QT interval (over 500 ms) on electrocardiography (ECG), a toxicity predicted in preclinical animal studies. These events occurred without any clinical correlate or symptoms. This dose-limiting toxicity was observed in Arm A at dose levels of 226 mg q12h ($n = 1$) and 339 mg q12h ($n = 2$), and in Arm B at 113 mg q12h ($n = 1$) and 226 mg q12h ($n = 2$). The incidence of grade 3 QT prolongation was 10% at dose levels either at or below MTD.

A total of 67 treated patients (99%) had an adverse event during treatment with revumenib whereas 53 patients (78%) had a treatment-related adverse event (TRAE) of any grade (Table 1). The most frequent treatment-emergent adverse events (TEAEs) of any grade, irrespective of a relationship to revumenib, were prolongation of the QT interval (56%), nausea (50%), vomiting (40%) and febrile neutropenia (31%); a full listing of these adverse events in at least 20% of the patients is given in Table 1. The most frequent

TEAEs of grade 3 or higher were febrile neutropenia (31%), thrombocytopenia (19%) and sepsis (18%) (Extended Data Table 2). The most frequent TRAE was prolongation of the QT interval, in 36 patients (53%). Eleven patients (16%) had grade 3 or 4 TRAEs (Extended Data Table 2). The most frequent grade 3 TRAE was prolongation of the QT interval, in nine patients (13%). There were no ventricular arrythmias, and all episodes of grade 3 QT prolongation resolved either before the next dose or with protocol-directed dose hold. All patients with grade 3 QT prolongation were able to resume dosing at the next-lower dose level. Other grade 3 TRAEs included diarrhoea (3%), hypercalcaemia (2%), tumour lysis syndrome (2%), anaemia (3%), asthenia (2%), fatigue (2%), neutropenia (2%), thrombocytopenia (2%), and hypokalaemia (2%). Grade 4 TRAEs included neutropenia (2%) and thrombocytopenia (2%), with one event each. No patients permanently discontinued revumenib due to TRAEs, and no TRAEs led to death.

Table 1 Any-grade treatment-related and TEAEs, regardless of causality

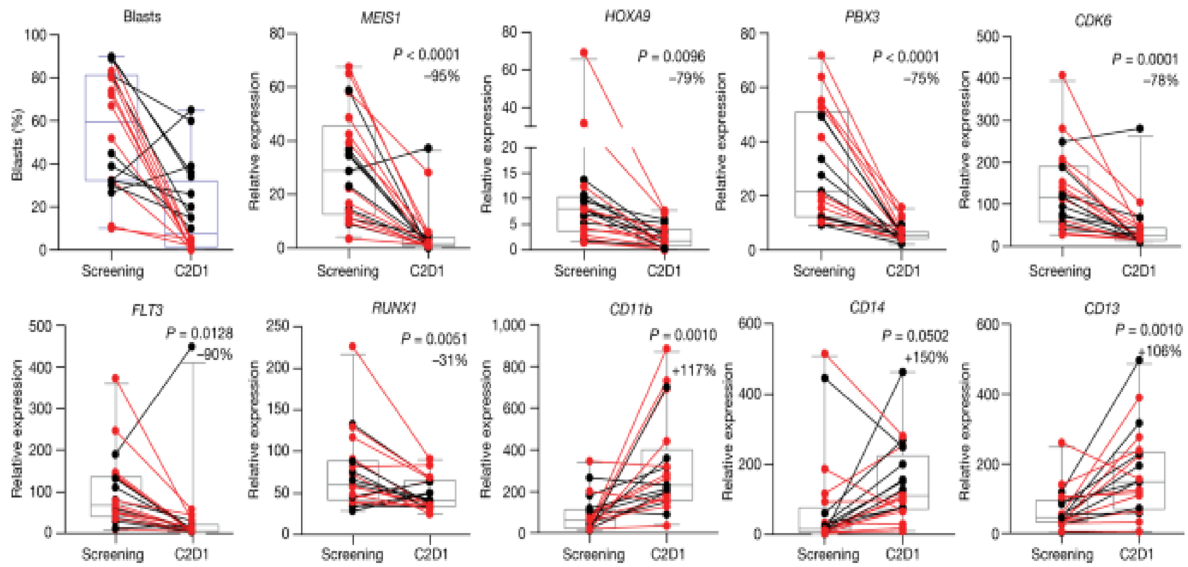
A common occurrence in antileukaemia therapies that induce myeloid differentiation is the development of a clinically apparent differentiation syndrome. Differentiation syndrome is caused by cytokine alterations associated with haematopoietic differentiation, with common manifestations including fever, arthralgias, leukocytosis, pleural or pericardial effusions and respiratory or renal failure in severe cases[14,15]. This had been mostly reported in patients with myeloid leukaemias treated with all-trans retinoic acid, arsenic trioxide and isocitrate dehydrogenase inhibitors[14,15,16]. In this study, differentiation syndrome was reported in 11 patients (16%), all grade 2. All cases of differentiation syndrome resolved following treatment with corticosteroids, with the addition of hydroxyurea for associated leukocytosis seen in five of the 11 patients (Extended Data Figs. 4 and 5). One patient missed one day of dosing, but otherwise treatment with revumenib was uninterrupted because of differentiation syndrome. The median time of onset of differentiation syndrome was 18 days (range, 5–41) with manifestations including bone pain or arthralgia, pericardial effusion, pleural effusion, pulmonary infiltrates, increase in creatinine, rash, oedema and pyrexia.

Pharmacokinetic studies showed that dose-proportional exposure was achieved in both Arm A (no strong CYP3A4 inhibitor) and Arm B (with a strong CYP3A4 inhibitor). Steady-state levels were achieved in approximately 48 h, with no evidence of drug accumulation (Extended Data Fig. 6).

Following a review of prespecified protocol criteria on safety, tolerability and pharmacokinetic data, the revumenib doses of 226 mg q12h and 276 mg q12h in Arm A and 113 mg q12h and 163 mg q12h in Arm B met the prespecified criteria for RP2D.

The pharmacodynamic effects of revumenib were assessed through transcriptional changes following treatment using RNA-seq of bone marrow cells (Fig. 1). Consistent with the established mechanism of action, menin inhibition using revumenib resulted in downregulation of the critical leukaemogenic target genes *MEIS1,* homeobox A9 (*HOXA9*), pre-B-cell leukaemia transcription factor 3 (*PBX3*) and cyclin-dependent kinase 6 (*CDK6*), and an increase in expression of genes associated with differentiation, such as integrin alpha M (*CD11b*) and *CD14*. We also observed transcriptional suppression of fms-like tyrosine kinase 3 (*FLT3*) following treatment, a putative transcriptional target of MEIS1; *FLT3* is frequently mutated in multiple subtypes of AML, including *KMT2A*r AML (10%), and is particularly common in AML with mutated *NPM1* (60%)[8,17,18,19].

Fig. 1: Transcriptional changes following treatment with the menin inhibitor revumenib in patients with relapsed or refractory acute leukaemia with *KMT2A*r or mutated *NPM1*.

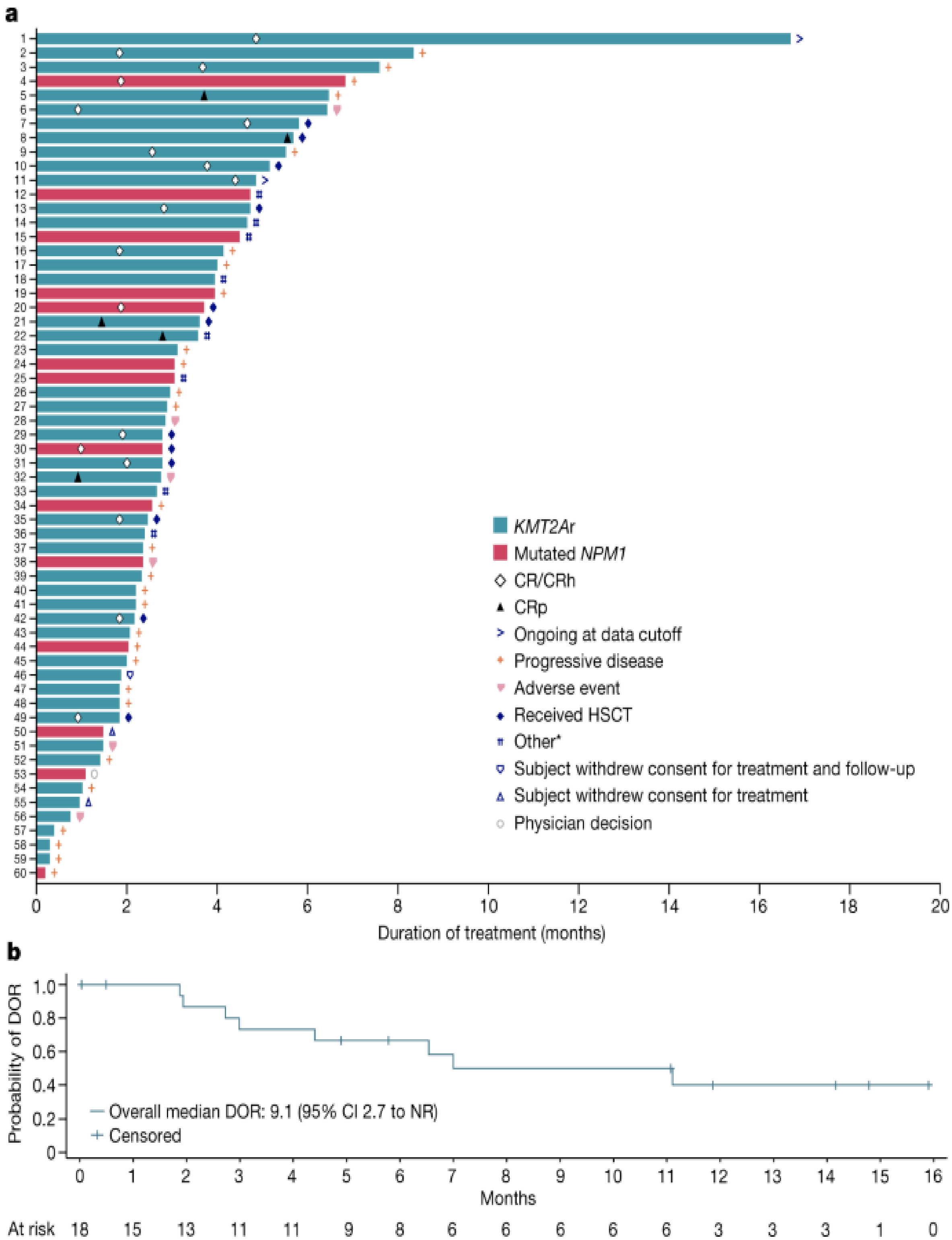


RNA-seq before and after treatment with revumenib, showing downregulation of critical leukaemogenic target genes *MEIS1*, *HOXA9* and *PBX3* and increase in expression of genes associated with differentiation (*CD11b*, *CD14*), with transcriptional suppression of *FLT3*, a putative transcriptional target of MEIS1. The change in bone marrow blast percentage following treatment is shown. Box plots represent median gene expression or median bone marrow blast percentage, and the 95% CI along with percentage change in gene expression following treatment. Responders are shown in red, nonresponders in black. Results were obtained using a paired *t*-test with a two-sided *P* value. Adjustments were not made for multiple comparisons. This analysis included a cohort of 21 evaluable patients. Revumenib was administered in continious 28-day cycles. C2D1, day 1 of treatment cycle 2.

The rate of complete remission or complete remission with partial haematologic recovery (CR/CRh) was 30% (95% confidence interval (CI) 18.8–43.2) in 18 of 60 evaluable patients with undetectable measurable residual disease (MRD), as assessed by multiparameter flow cytometry, in 14 of 18 patients (78%) who achieved CR/CRh (Fig. 2a and Table 2). The overall response rate (CR/CRh/complete remission with incomplete platelet recovery (CRp)/complete remission with incomplete haematologic recovery (CRi)/morphologic leukaemia-free state) was 53% (32 of 60 evaluable patients). The median time to CR/CRh was 1.9 months (range, 0.9–4.9). Although imaging assessment was not mandated on study, responses were notably seen in both bone marrow and extramedullary sites in two of six

evaluable patients with extramedullary leukaemia at enrolment, with representative imaging shown in Extended Data Fig. 7. In an exploratory descriptive analysis, we evaluated responses by both leukaemia lineage and the adult and paediatric (under 18 years of age) populations. Morphologic remissions were identified in 27 of 49 patients (55%) with AML (95% CI 40.2–69.3), in four of ten patients (40%) with ALL (95% CI 12.2–73.8) and in the one patient with a mixed-phenotype acute leukaemia. Morphologic remission was noted in four of eight paediatric patients (50%; 95% CI 15.7–84.3) and in 28 of 52 adults (54%; 95% CI 39.5–67.8).

Fig. 2: Characterization of remissions with the menin inhibitor revumenib in susceptible relapsed or refractory acute leukaemia subtypes.



a, Time to response, duration of treatment (censored at time of HSCT) and patient status by the cutoff date. *Other reasons for treatment discontinuation included no response, relapse, death and donor lymphocyte

infusion. **b**, Kaplan–Meier curve of duration of response (DOR) in patients with CR or CR/CRh without censoring at the time of an allogeneic stem cell transplant performed in 12 of 18 evaluable patients.

Table 2 Responses to treatment

In patients with *KMT2A*r who achieved morphologic remission following the first cycle of therapy with revumenib, 21 of 25 (84%) retained the detectable fusions causing *KMT2A*r during concomitant cytogenetic analyses. However, most patients later had clearance of these *KMT2A* rearrangements with subsequent cycles of therapy. The rate of complete cytogenetic response in patients with *KMT2A*r who achieved morphologic clearance of their myeloblasts was 16 of 25 (64%) (Extended Data Table 3); these typically occurred after morphologic remission and, in some cases, after MRD negativity as assessed by immunophenotyping via flow cytometry. The median time to achieving complete cytogenetic response was 1.9 months (range, 0.9–2.8). These findings are consistent with the mechanism of action of differentiation agents, for which response leads first to a phenotypic differentiation of cells such that they retain the genetic fusion but are without aberrant morphology or leukaemia surface marker expression (as detected by flow cytometry)[20,21]. Subsequently with additional cycles following phenotypic remission, fusions are no longer detected by fluorescence in situ hybridization, probably because of apoptosis of differentiated cells or replacement by normal haematopoiesis.

With a median follow-up of 11.9 months (95% CI 4.9–14.8) in patients who achieved CR/CRh, the median duration of response was 9.1 months (95% CI 2.7 to not reached (NR)) (Fig. 2b). Median overall survival in the efficacy population, regardless of remission status, was 7 months (95% CI 4.3–11.6) (Extended Data Fig. 8) and median follow-up of overall survival in the efficacy population was 14.3 months (95% CI 10.6–16.7). Twelve patients received an allogeneic stem cell transplant as consolidation following response to revumenib, with nine in remission at the time of the data cutoff, seven of whom have been in remission for over 6 months. The median duration of response in those who achieved CR/CRh with censoring at the time of allogeneic stem cell transplant was 4.4 months (95% CI 1.9–NR) (Extended Data Fig. 9).

Consistent with the preclinical hypothesis regarding the efficacy of menin inhibition in patients with *NPM1* mutations or *KMT2A*r, there were no responses in the eight patients enroled on study who lacked either a *KMT2A*r or *NPM1* mutation. In evaluable patients for whom a targeted sequencing panel was performed at baseline, no specific *KMT2A* translocation or comutation was clearly associated with response in a limited analysis (Extended Data Fig. 10). Notably among evaluable patients with *KMT2A*r or mutated *NPM1*, 14 had a comutation in *FLT3*, of whom 13 had been treated with a FLT3 inhibitor before enrolment. Among the 11 patients with comutated *NPM1* and *FLT3*, three responded whereas two of the three patients with *KMT2A*r and concomitant *FLT3* mutations achieved a response. All responders had previously received a FLT3 inhibitor.

Discussion

The treatment of patients who have relapsed or refractory acute leukaemia with *KMT2A*r is challenging. The rate of CR or complete remission with incomplete count recovery in relapsed *KMT2A*r AML following two previous lines of therapy is 9% (ref. [18]). Despite notable progress in the treatment of childhood acute leukaemia, infant *KMT2A*r acute leukaemias have remained a therapeutic challenge with high rates of resistance to multi-agent chemotherapy[22,23]. Mutations in *NPM1* are the most common alterations in AML, and targeting of vulnerabilities associated with this subset expand the reach of targeted therapy to the largest portion of patients with this disease. The menin inhibitor revumenib has the potential to address these unmet needs. In this first-in-human clinical trial, we provide clinical evidence of the effectiveness of menin inhibition with an oral targeted therapy, which is the first epigenetic therapy that evicts protein complexes from chromatin, leading to remissions in patients with acute leukaemia. We found an encouraging clinical benefit, with deep molecular remissions and minimal toxicities, in a heavily pretreated population of both children and adults with advanced acute leukaemia. However, recent data suggest that clinical resistance to targeted menin inhibition is mediated through the acquisition of mutations in menin that prevent inhibitor binding and can lead to clinical relapse[24]. The emergence of resistance mutations highlights the critical role of the menin–KMT2A interaction in the pathogenesis of *KMT2A*r and *NPM1*-mutant acute leukaemia.

Because revumenib disrupts the effect of epigenetic regulators in leukaemia that are dependent on the menin–KMT2A interaction, this leads to abrogation of aberrant gene expression and removal of the haematopoietic differentiation block[12,13]. In this study, we show evidence of this mechanism of action with differentiation of leukaemia cells, persistence of cytogenetic abnormalities at the time of morphologic response and differentiation syndrome associated with menin inhibition in some patients, which resolved following appropriate therapy.

Other leukaemia genotypes have recently been identified as susceptible to menin inhibition, such as AML with rearrangement of the nucleoporin 98 gene (*NUP98*), a common and adverse genotype among children with relapsed and refractory disease[25]. Another example is the rare occurrence of AML with translocations involving the meningioma-1 gene (MN1)[26]. Therefore, menin inhibitors have the potential to reach a larger subset of acute leukaemia with similar dependency on the menin–KMT2A interaction, which could be identified using precision approaches testing characteristic gene expression.

In conclusion, in children and adults with highly refractory acute leukaemia with *KMT2A*r or *NPM1* mutation, menin inhibition with revumenib monotherapy was associated with promising antileukaemic activity leading to deep and sustained remission.

Methods

Study design and oversights

We conducted a phase 1, multicentre, open-label, dose-escalation study (ClinicalTrials.gov, NCT04065399). Revumenib was administered orally in either capsule or liquid formulation, q12h, in 28-day continuous cycles with dose adjustment by body surface area for patients weighing under 40 kg (Extended Data Fig. 3). We employed a rolling-6 design, an algorithm-based extension of the 3+3 design that allows for concurrent accrual of two to six patients onto a dose level, thereby shortening the duration of phase 1 without increasing the risk of toxicity[27]. Therefore, the number of patients enrolled followed this dose-escalation design

without an absolute prior estimation of sample size. The study included two parallel dose-escalation cohorts, for patients either not taking (Arm A) or taking (Arm B) strong CYP3A4 inhibitors. Dose expansion occurred at safe and efficacious doses. Dose-limiting toxic effects were defined as nonhaematologic toxic effects of grade 3 or higher during cycle 1, or as haematologic toxicities directly attributed to the study drug rather than to the underlying disease.

The study was conducted in accordance with the principles of the Declaration of Helsinki and the International Council for Harmonization Good Clinical Practice. The protocol and amendments were approved by the institutional review board or ethics committee at The University of Texas MD Anderson Cancer Center, City of Hope, Washington University School of Medicine in St. Louis, Dana-Farber Cancer Institute, Winship Cancer Institute, Emory University School of Medicine, University of Chicago, Florida Cancer Specialists/Sarah Cannon Research Institute, University of Iowa and Memorial Sloan Kettering Cancer Center. All patients provided written informed consent. This study was designed by the sponsor (Syndax Pharmaceuticals) in collaboration with the investigators. The data were collected by the investigators and their research staff and were analysed by the sponsor and authors. Drafts of the manuscript were written by the first and last authors and revised in collaboration with all authors and the sponsor, all of whom vouch for the completeness and accuracy of the data and analyses, and for the adherence of the study to the protocol. Assistance in manuscript preparation for submission was provided by a professional service and paid for by the sponsor.

Patients

An early amendment to the protocol allowed patients aged 30 days and older to be enroled on this study; additionally, it restricted the eligibility criteria from any relapsed or refractory acute leukaemia to only those with *KMT2A*r or mutated *NPM1* in Arms A and B of the study. Mutational status was assessed at each site.

Study assessments

The primary objectives of the phase 1 portion of this study were to assess safety, the MTD or, if different, RP2D and to characterize the pharmacokinetic parameters of revumenib in arms based on concomitant CYP3A inhibitors. Exploratory objectives included assessment of antileukaemic activity in the efficacy population, which consisted of patients with *KMT2A*r or mutated *NPM1*. All study assessments and analyses included both paediatric and adult patients.

Adverse events were graded with use of the Common Terminology Criteria for Adverse Events, v.5.0. Clinical efficacy was assessed by the investigators with use of a modified version of the 2017 European LeukemiaNet response criteria, to additionally include CRh[28]. Guidelines for managing prolongation of the QT interval included electrolyte repletion and adjustment of revumenib dose. Investigators were encouraged to administer corticosteroids following suspicion of differentiation syndrome based on guidelines included in the protocol. Measurable residual disease assessment was performed in participating institutions using multicolour flow cytometry or PCR[29,30,31]. In an exploratory mutational analysis we used the ArcherDx VariantPlex Myeloid Panel, a 75-gene, next-generation targeted sequencing product examining genes frequently mutated in myeloid and lymphoid malignancies. Powered by Anchored Multiple PCR chemistry, the panel enables deep strand-specific amplification of unique molecular barcoded DNA fragments for sequencing. Archer Analysis software was used for the analysis and interpretation of sequencing data in the detection of single-nucleotide variants and indels. Variant calling was performed using Invitae's Comprehensive Targeted Mutation File (v.1.6), and the Vision variant caller to report well-classified variants.

Statistical analysis

All patients who received at least one dose of revumenib were included in the safety analysis, whereas only those with *KMT2A*r or mutated *NPM1* were included in the efficacy analysis. Time-to-event end points were estimated using the Kaplan–Meier method. Descriptive statistics were used for other clinical, laboratory and pharmacokinetic variables. Clinical data were captured in Medidata Classic Rave 2021.2.0, and data analyses were

performed using SAS v.9.4 and GraphPad Prism v.8.0. Changes in gene expression were analysed using a paired *t*-test.

RP2D was determined by the safety review committee based on review of pharmacokinetics, safety and tolerability data among evaluable patients.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Syndax Pharmaceuticals (Syndax) is committed to providing qualified scientific researchers access to anonymized data and clinical study reports from the company's clinical trials for the purpose of conducting legitimate scientific research. Syndax is also obligated to protect the rights and privacy of trial participants and, as such, has a procedure in place for evaluating and fulfilling requests for sharing company clinical trial data with qualified external scientific researchers. Following submission of a request (to datarequest@syndax.com), Syndax will provide an outline of the process and requirements for submitting a data request. Feasible requests will be reviewed by a committee of Syndax subject matter experts to assess the scientific validity of the request and the qualifications of the requestors. In line with data privacy legislation, submitters of approved requests must enter into a standard data-sharing agreement with Syndax before Syndax may grant data access. Data will be made available for request after product approval in the United States and European Union, or after product development is discontinued. There are circumstances that may prevent Syndax from sharing requested data, including country- or region-specific regulations. If Syndax declines the request, it will communicate the decision to the investigator. Access to genetic or exploratory biomarker data requires a detailed statistical analysis plan that is collaboratively developed by the requester and Syndax subject matter experts.

References

1. Lo-Coco, F. et al. Retinoic acid and arsenic trioxide for acute promyelocytic leukemia. *N. Engl. J. Med.* **369**, 111–121 (2013).

2. DiNardo, C. D. et al. Durable remissions with ivosidenib in IDH1-mutated relapsed or refractory AML. *N. Engl. J. Med.* **378**, 2386–2398 (2018).

3. Stein, E. M. et al. Enasidenib in mutant IDH2 relapsed or refractory acute myeloid leukemia. *Blood* **130**, 722–731 (2017).

4. Yokoyama, A. et al. The menin tumor suppressor protein is an essential oncogenic cofactor for MLL-associated leukemogenesis. *Cell* **123**, 207–218 (2005).

5. Grembecka, J. et al. Menin-MLL inhibitors reverse oncogenic activity of MLL fusion proteins in leukemia. *Nat. Chem. Biol.* **8**, 277–284 (2012).

6. Kühn, M. W. et al. Targeting chromatin regulators inhibits leukemogenic gene expression in NPM1 mutant leukemia. *Cancer Discov.* **6**, 1166–1181 (2016).

7. Krivtsov, A. V. & Armstrong, S. A. MLL translocations, histone modifications and leukaemia stem-cell development. *Nat. Rev. Cancer* **7**, 823–833 (2007).

8. Papaemmanuil, E. et al. Genomic classification and prognosis in acute myeloid leukemia. *N. Engl. J. Med.* **374**, 2209–2221 (2016).

9. Issa, G. C. et al. Therapeutic implications of menin inhibition in acute leukemias. *Leukemia* **35**, 2482–2495 (2021).

10. Thorsteinsdottir, U., Kroon, E., Jerome, L., Blasi, F. & Sauvageau, G. Defining roles for HOX and MEIS1 genes in induction of acute myeloid leukemia. *Mol. Cell. Biol.* **21**, 224–234 (2001).

11. Yu, B. D., Hess, J. L., Horning, S. E., Brown, G. A. J. & Korsmeyer, S. J. Altered Hox expression and segmental identity in Mll-mutant mice. *Nature* **378**, 505–508 (1995).

12. Uckelmann, H. J. et al. Therapeutic targeting of preleukemia cells in a mouse model of *NPM1* mutant acute myeloid leukemia. *Science* **367**, 586–590 (2020).

13. Krivtsov, A. V. et al. A menin-MLL inhibitor induces specific chromatin changes and eradicates disease in models of MLL-rearranged leukemia. *Cancer Cell* **36**, 660–673 (2019).

14. Sanz, M. A. & Montesinos, P. How we prevent and treat differentiation syndrome in patients with acute promyelocytic leukemia. *Blood* **123**, 2777–2782 (2014).

15. Fathi, A. T. et al. Differentiation syndrome associated with enasidenib, a selective inhibitor of mutant isocitrate dehydrogenase 2: analysis of a phase 1/2 study. *JAMA Oncol.* **4**, 1106–1110 (2018).

16. Norsworthy, K. J. et al. Differentiation syndrome with ivosidenib and enasidenib treatment in patients with relapsed or refractory IDH-mutated AML: a U.S. Food and Drug Administration systematic analysis. *Clin. Cancer Res.* **26**, 4280–4288 (2020).

17. Dzama, M. M. et al. Synergistic targeting of FLT3 mutations in AML via combined menin-MLL and FLT3 inhibition. *Blood* **136**, 2442–2456 (2020).

18. Issa, G. C. et al. Predictors of outcomes in adults with acute myeloid leukemia and KMT2A rearrangements. *Blood Cancer J.* **11**, 162 (2021).

19. Bill, M. et al. Mutational landscape and clinical outcome of patients with de novo acute myeloid leukemia and rearrangements involving 11q23/KMT2A. *Proc. Natl Acad. Sci. USA* **117**, 26340–26346 (2020).

20. Grimwade, D. et al. Prospective minimal residual disease monitoring to predict relapse of acute promyelocytic leukemia and to direct pre-emptive arsenic trioxide therapy. *J. Clin. Oncol.* **27**, 3650–3658 (2009).

21. Andreeff, M. et al. Hexamethylene bisacetamide in myelodysplastic syndrome and acute myelogenous leukemia: a phase II clinical trial

with a differentiation-inducing agent. *Blood* **80**, 2604–2609 (1992).

22. Hilden, J. M. et al. Analysis of prognostic factors of acute lymphoblastic leukemia in infants: report on CCG 1953 from the Children's Oncology Group. *Blood* **108**, 441–451 (2006).

23. Pieters, R. et al. A treatment protocol for infants younger than 1 year with acute lymphoblastic leukaemia (Interfant-99): an observational study and a multicentre randomised trial. *Lancet* **370**, 240–250 (2007).

24. Perner, F. et. al. *MEN1* mutations mediate clinical resistance to menin inhibition. *Nature* https://doi.org/10.1038/s41586-023-05755-9 (2023).

25. Heikamp, E. B. et al. The menin-MLL1 interaction is a molecular dependency in *NUP98*-rearranged AML. *Blood* **139**, 894–906 (2021).

26. Libbrecht, C. et al. Menin is necessary for long term maintenance of meningioma-1 driven leukemia. *Leukemia* **35**, 1405–1417 (2021).

27. Skolnik, J. M. et al. Shortening the timeline of pediatric phase I trials: the rolling six design. *J. Clin. Oncol.* **26**, 190–195 (2008).

28. Döhner, H. et al. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood* **129**, 424–447 (2017).

29. Buccisano, F. et al. The kinetics of reduction of minimal residual disease impacts on duration of response and survival of patients with acute myeloid leukemia. *Leukemia* **20**, 1783–1789 (2006).

30. Terwijn, M. et al. High prognostic impact of flow cytometric minimal residual disease detection in acute myeloid leukemia: data from the HOVON/SAKK AML 42A study. *J. Clin. Oncol.* **31**, 3889–3897 (2013).

31. Freeman, S. D. et al. Prognostic relevance of treatment response measured by flow cytometric residual disease detection in older patients with acute myeloid leukemia. *J. Clin. Oncol.* **31**, 4123–4131 (2013).

Acknowledgements

This study was funded by Syndax Pharmaceuticals. We thank the patients, their families and caregivers for participating in the study, in addition to all of the participating investigators and research staff. Bioinformatics support was provided by Fios Genomics. G.C.I. was additionally supported by the K12 Paul Calabresi Clinical Scholarship Award (no. NIH/NCI K12 CA088084) and E.M.S. by NIH/NCI Cancer Center Support Grant no. P30 CA008748 to Memorial Sloan Kettering Cancer Center.

Author information

Authors and Affiliations

1. Department of Leukemia, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

   Ghayas C. Issa, Branko Cuglievan & Hagop Kantarjian

2. City of Hope, Duarte, CA, USA

   Ibrahim Aldoss

3. Washington University School of Medicine in St. Louis, St. Louis, MO, USA

   John DiPersio & Shalini Shenoy

4. Dana-Farber Cancer Institute, Boston, MA, USA

   Richard Stone, Scott A. Armstrong, Florian Perner & Jennifer A. Perry

5. Winship Cancer Institute, Emory University School of Medicine, Atlanta, GA, USA

   Martha Arellano

6. University of Chicago, Chicago, IL, USA

Michael J. Thirman

7. Florida Cancer Specialists/Sarah Cannon Research Institute, Sarasota, FL, USA

Manish R. Patel

8. University of Iowa, Iowa City, IA, USA

David S. Dickens

9. Memorial Sloan Kettering Cancer Center, New York, NY, USA

Neerav Shukla & Eytan M. Stein

10. Greifswald University Medical Center, Greifswald, Germany

Florian Perner

11. Syndax Pharmaceuticals, Waltham, MA, USA

Galit Rosen, Rebecca G. Bagley, Michael L. Meyers, Peter Ordentlich, Yu Gu, Vinit Kumar, Steven Smith & Gerard M. McGeehan

Contributions

G.C.I., E.M.S., G.R., M.L. M. and Y.G. conceptualized and designed the study. G.C.I. and E.M.S. prepared the first draft of the manuscript, provided critical review and revisions of the drafts and approved the decision to submit for publication. F.P., S.A.A. and J.A.P. designed, performed and analysed preclinical experiments using patient-derived xenografts. G.C.I., I.A., J.D.P., B.C., R.S., M.A., M.J.T., M.R.P., D.S.D., S.Shenoy, N.S., H.K. and E.M.S. enroled and treated patients. G.C.I., I.A., J.D.P., B.C., R.S., M.A., M.J.T., M.R.P., D.S.D., S.Shenoy, N.S., S.A.A., F.P., J.A.P., G.R., R.G.B., M.L.M., P.O., Y.G., V.K., S.Smith, G.M.M. and E.M.S. collected data, analysed and interpreted the results and reviewed and edited manuscript drafts. S.Smith participated in conception and design analysis and interpretation of results. M.R.P. and S.A.A. participated in trial conduct and clinical trial discussions. G.M.M. conceptualized and designed the

study, analysed data and interpreted results for preclinical studies. V.K. collected data and analysed and interpreted results for pharmacodynamic studies. All authors have read and agreed on the content of the manuscript.

Corresponding authors

Correspondence to Ghayas C. Issa or Eytan M. Stein.

Ethics declarations

Competing interests

G.C.I. received consultancy or advisory role fees from Novartis, Kura Oncology and NuProbe and received research funding from Celgene, Novartis, Kura Oncology, Syndax Pharmaceuticals, Merck, Cullinan Oncology and NuProbe. I.A. received consultancy or advisory role fees from Amgen, Pfizer, Jazz, AbbVie and Agios, research funding from AbbVie and Macrogenics and honoraria from Amgen, Pfizer, Jazz, AbbVie and Agios. J.D.P. has a consultancy role with Incyte and RiverVest Venture Partners, has served as a board member or advisory committee member for RiverVest Venture Partners, Magenta, hC Bioscience, Inc. and WUGEN, has received research funding from NeoImmune Tech, Macrogenics, Incyte, Bioline Rx and WUGEN and holds patents or pending patents for UCART7 for treatment of T-ALL, VLA-4 inhibitors for stem cell mobilization and NT-I7 for CART expansion. R.S. has served on the steering committee of AbbVie and advisory boards of AbbVie, AvenCell, CTI Pharma, Kura One, Genentech, Actinium, Arog, BMS, Boston Pharmaceuticals, GSK, Janssen, Jazz, Novartis, Syros, Takeda, Elevate Bio, Syndax Pharmaceuticals, Gemoab, BerGenBio, Foghorn Tera, Aprea, Innate, Actinium and OncoNova; served as DSMB for Aptevo, Epizyme, Takeda and Syntrix/ACI Clinical; on the focus group of BerGenBio; and on AML Expert Council of GSK and Grand Rounds of Jazz Pharmaceuticals. M.A. has served on the advisory boards of Kite Pharma and Syndax Pharmaceuticals. M.J.T. has a consulting or advisory role with AbbVie and CVS, has an expert testimony role with Apotex and received research funding from AbbVie, Gilead Sciences, Janssen, Merck, Pharmacyclics, Syndax Pharmaceuticals, TG Therapeutics and Tolero. M.R.P. served in a leadership role with ION

Pharma; received honoraria from Pfizer, Pharmacyclics, Bayer, Janssen Oncology, Genentech and Adaptive Biotechnologies; has a consulting or advisory role with Pharmacyclics/Janssen and Pfizer/EMD Serono; served on the Speakers' Bureau of Exelixis, Genentech/Roche, Taiho Pharmaceutical and Celgene; and received research funding from Acerta Pharma, ADC Therapeutics, Agenus, Aileron Therapeutics, AstraZeneca, BioNTech AG, Boehringer Ingelheim, Celgene, Checkpoint Therapeutics, CicloMed, Clovis Oncology, Cyteir Therapeutics, Daiichi Sankyo, Lilly, EMD Serono, Evelo Therapeutics, FORMA Therapeutics, Genentech/Roche, Gilead Sciences, GlaxoSmithKline, H3 Biomedicine, Hengrui Therapeutics, Hutchison MediPharma, Ignyta, Incyte, Jacobio, Janssen, Klus Pharma, Kymab, Loxo, LSK Biopartners, Lycera, Macrogenics, Merck, Millennium, Mirati Therapeutics, Moderna Therapeutics, Pfizer, Placon, Portola Pharmaceuticals, Prelude Therapeutics, Ribon Therapeutics, Seven and Eight Biopharmaceuticals, Syndax Pharmaceuticals, Taiho Pharmaceutical, Takeda, Tesaro, TopAlliance BioSciences, Inc., Vigeo, ORIC, Artios, IgM Biosciences, Puretech, BioTheryX, Black Diamond Therapeutics, IgM Biosciences, NGM Biopharmaceuticals, Nurix, PureTech, Relay Therapeutics, Samumed, Silicon Therapeutics, TeneoBio, Treadwell Therapeutics, Zymeworks, Olema, Adagene, Astellas, NGM, Accutar Biotech, TeneoBio, Novartis, Compugen, Black Diamond Therapeutics, MabSpace Biosciences, Immunogen and Blueprint Pharmaceuticals. D.S.D. has a consulting or advisory role with Tempus, Inc. S. Shenoy has a consulting or advisory role with Artio, BMS and Takaeda. H.K. received honoraria/advisory board/consulting fees from AbbVie, Amgen, Amphista, Ascentage, Astellas, Biologix, Curis, Ipsen Biopharmaceuticals, KAHR Medical, Labcorp, Novartis, Pfizer, Shenzhen Target Rx, Stemline and Takeda; and received research funding from AbbVie, Amgen, Ascentage, BMS, Daiichi Sankyo, Immunogen, Jazz, and Novartis. S.A.A. received stock or other ownership from Neomorph, Inc., C4 Therapeutics, Cyteir Therapeutics, Accent Therapeutics and Mana Therapeutics; has a consulting or advisory role with Neomorph, Inc., C4 Therapeutics, Cyteir Therapeutics, Accent Therapeutics, Mana Therapeutics and Twentyeight-Seven Therapeutics; received research funding from Syndax Pharmaceuticals and Janssen; and holds patents, royalties and other intellectual property for MENIN inhibition in NPM1 AML: WO/2017/132398A1. G.R. is a former employee of Syndax

Pharmaceuticals and a current employee of Boston Pharmaceuticals. R.G.B. is an employee of Syndax Pharmaceuticals and has stock or other ownership at Syndax Pharmaceuticals. M.L.M., P.O. and G.M.M. are employees of Syndax Pharmaceuticals. M.L.M. has a consulting or advisory role at Nuvalent, holds patents, royalties and other intellectual property at Syndax Pharmaceuticals and Nuvalent and has stock or other ownership at Syndax Pharmaceuticals and Johnson & Johnson. P.O. has a consulting or advisory role at Patrys and Twentyeight-Seven Therapeutics, holds patents, royalties and other intellectual property at Syndax Pharmaceuticals and has stock or other ownership at Syndax Pharmaceuticals. G.M.M. has a consulting or advisory role at Syndax Pharmaceuticals, holds patents, royalties and other intellectual property at Syndax Pharmaceuticals and has stock or other ownership at Syndax Pharmaceuticals. Y.G. is an employee of Syndax Pharmaceuticals and has stock or other ownership at Syndax Pharmaceuticals and AstraZeneca. S. Smith has a consultancy or advisory role with Syndax Pharmaceuticals. E.M.S. has a consulting or advisory role with Gilead, CTI Biopharma, Epizyme, AbbVie, Pinotbio, Neoleukin Genesis, Genentech, Jazz, Novartis, Celgene, Calithera, Takeda, Janssen, BMS, Kronos, Kura, Auron, Syndax Pharmaceuticals, Servier, Agios and Remix and received research funding from Biotheryx, Agios, Servier, Eisai, BMS, Bayer, Syndax, Syros and Loxo. B.C., F.P., J.A.P., N.S. and V.K. declare no competing interests.

Peer review

Peer review information

Nature thanks Jorge Cortes, Eunice Wang and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Revumenib suppresses both *KMT2A*r and *NPM1*-mutant AML in preclinical models of leukaemia.

NSG mice were engrafted with the *KMT2A*r cell line MOLM-13 (**a**; n = 10) or patient derived xenografts (PDX) harbouring either a *KMT2A*r (**d**; n = 5) or *NPM1* mutation (**g**; n = 5). Mice were treated for 28 days with revumenib, which was formulated in chow, across a dose range of 0.025% to 0.2% (**a**) or at 0.1% fixed dose (**d**, **g**). Leukaemic burden (CD45+) was assessed at end of treatment (**b**; 0.025%: n = 3; 0.05-0.2%: n = 5, data represent mean ± SEM) or throughout the study (**d**, **g;** data represent mean ± SEM). For MOLM-13 engrafted mice, revumenib showed clear dose-dependent exposure (**c**; n = 3 with 3 individual measurements per timepoint and dose, data represent mean ± SD), which translated into dose-responsive effect on survival benefit (**a**) and leukaemic burden at end of treatment (**b**). Similarly, revumenib treatment of the PDX models led to significant suppression of leukaemic burden and significant survival benefit in each (**d**, **g**). *P*-values were determined by log-rank (Mantel–Cox) tests. Adjustments were not made for multiple comparisons. Revumenib treatment also led to broad changes in the transcriptional program (**f**, **i**; n = 3 per treatment group; GEO accession number for RNAseq data, GSE216730) with GSEA results consistent with previously reported signatures (**e**, **h**; n = 3 per treatment group). GSEA, gene set enrichment analysis; NES, normalized enrichment score; NSG, NOD scid gamma; PDX, patient-derived xenograft; SD, standard deviation; SEM, standard error of the mean.

Extended Data Fig. 2 CONSORT diagram and patient disposition on trial.

The CONSORT diagram shows the number of patients from Arms A and B who discontinued treatment and lists the reasons for treatment and study discontinuation.

Extended Data Fig. 3 Dose escalation schema.

The revumenib dose was adjusted based on the body surface area (BSA) for patients weighing less than 40 kg as indicated for each corresponding dose

level shown in parentheses. CYP3A4i, cytochrome P450 3A4 inhibitor; q12h, every 12 h; R/R, relapsed or refractory.

Extended Data Fig. 4 Morphologic evidence of myeloid differentiation.

Photomicrographs of bone marrow biopsies demonstrating morphologic evidence of myeloid differentiation in a patient who achieved complete remission; magnification is 40x.

Extended Data Fig. 5 Changes in peripheral blood during differentiation syndrome are associated with the menin inhibitor revumenib.

Example from a 71-year-old patient with *KMT2A*r AML relapsed after an allogeneic stem cell transplant, who received revumenib at 339 mg PO q12h (Arm A), and achieved CRh, MRD negative remission. Differentiation syndrome manifested as chest pain with a small pericardial effusion, and a possible prodrome of neck pain likely related to expansion of cervical nodes, all resolved promptly with initiation of steroids followed by tapering doses. Hydroxyurea was used to control leukocytosis. AML, acute myeloid leukaemia; ANC, absolute neutrophil count; CRh, complete remission with partial haematologic recovery; MRD, minimal or measurable residual disease; PB, peripheral blood; PO, by mouth; q12h, every 12 h; WBC, white blood cell.

Extended Data Fig. 6 Dose proportional exposure was achieved across both arms.

The half-life of revumenib in Arm A (without strong CYP3A4 inhibitors) was approximately 3 h at the cycle 1 day 8 assessment of the 276-mg q12h dose level and was approximately 8 h at the same assessment in Arm B (with a strong CYP3A4 inhibitor) at the 163-mg q12h dose level. Data represent mean ± SD. Data cutoff date for pharmacokinetic analysis was July 11, 2022. CYP3A4, cytochrome P450 3A4; q12h, every 12 h; SD, standard deviation.

Extended Data Fig. 7 Response in extra-medullary disease.

A. PET scan at baseline and after 2 cycles of treatment. **B**. Computed tomographic scans of target lesions and spleen from a 19-year-old with relapsed *KMT2A*r AML, previously treated with three prior lines of therapy including 2 allogeneic stem cell transplants and local radiation to the spleen and abdominal nodes, received revumenib at 276 mg PO q12h (Arm A), achieved CRh, MRD negative remission with resolution of extramedullary disease in abdominal nodes and spleen. AML, acute myeloid leukaemia; AP, anterior-posterior; CRh, complete remission with partial haematologic recovery; D, day; FDG, fluorodeoxyglucose; MRD, minimal or measurable residual disease; PET, positron emission tomography; PO, by mouth; q12h, every 12 h; SUV, standardized uptake value.

Extended Data Fig. 8 Overall survival in patients with *KMT2A*r or mutated *NPM1*.

OS, overall survival.

Extended Data Fig. 9 Kaplan-Meier curve of duration of response in patients with complete remission or complete remission with partial haematologic recovery with censoring at time of the allogeneic stem cell transplant.

DOR, duration of response; NR, not reached.

Extended Data Fig. 10 Mutation analysis and response.

This cohort included 37 patients evaluable for this analysis. CR, complete remission; CRh, complete remission with partial haematologic recovery; CRp, complete remission with incomplete platelet recovery; MLFS, morphologic leukaemia-free state; NR, not reached; PD, progressive dVAF, variant allele frequency.

Extended Data Table 1 Baseline characteristics

Extended Data Table 2 Grade 3 or higher treatment-related and treatment-emergent adverse events
Extended Data Table 3 Measurable residual disease and cytogenetic responses

Supplementary information

Reporting Summary

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Issa, G.C., Aldoss, I., DiPersio, J. *et al.* The menin inhibitor revumenib in *KMT2A*-rearranged or *NPM1*-mutant leukaemia. *Nature* **615**, 920–924 (2023). https://doi.org/10.1038/s41586-023-05812-3

- Received: 06 September 2022

- Accepted: 08 February 2023
- Published: 15 March 2023
- Issue Date: 30 March 2023
- DOI: https://doi.org/10.1038/s41586-023-05812-3

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

[Inhibition of the protein menin shows early promise in leukaemia]

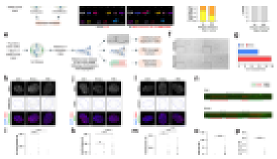
Nature Clinical Briefing 15 Mar 2023

[Menin–KMT2A interaction inhibitor shows promise]

- David Killock

Nature Reviews Clinical Oncology Research Highlight 29 Mar 2023

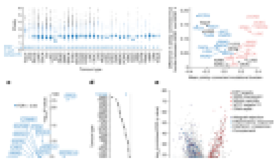
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05812-3



Short-term molecular consequences of chromosome mis-segregation for genome stability

11 March 2023

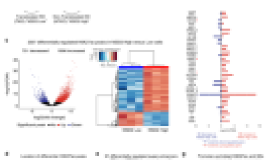
Lorenza Garribba, Giuseppina De Feudis, … Stefano Santaguida



Whole-genome doubling confers unique genetic vulnerabilities on tumour cells

27 January 2021

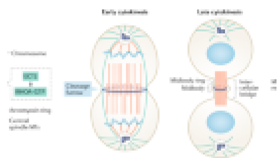
Ryan J. Quinton, Amanda DiDomizio, … Neil J. Ganem



NSD2 overexpression drives clustered chromatin and transcriptional changes in a subset of insulated domains

24 October 2019

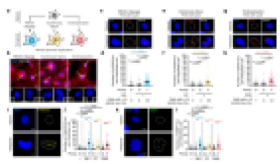
Priscillia Lhoumaud, Sana Badri, … Jane A. Skok



[Cytokinesis defects and cancer](#)

06 December 2018

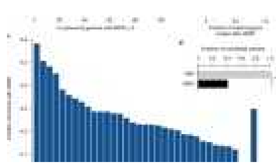
Susanne M. A. Lens & René H. Medema



[Genetic instability from a single S phase after whole-genome duplication](#)

30 March 2022

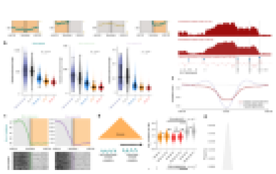
Simon Gemble, René Wardenaar, … Renata Basto



[Genome doubling shapes the evolution and prognosis of advanced cancers](#)

16 July 2018

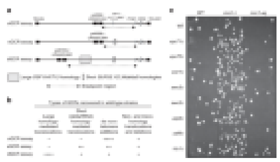
Craig M. Bielski, Ahmet Zehir, … Barry S. Taylor



Somatic mutation distributions in cancer genomes vary with three-dimensional chromatin structure

05 October 2020

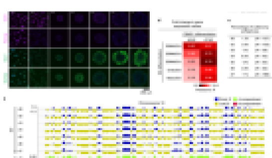
Kadir C. Akdemir, Victoria T. Le, … P. Andrew Futreal



The Swr1 chromatin-remodeling complex prevents genome instability induced by replication fork progression defects

11 September 2018

Anjana Srivatsan, Bin-Zhong Li, … Richard D. Kolodner



Single-cell DNA replication profiling identifies spatiotemporal developmental dynamics of chromosome organization

12 August 2019

Hisashi Miura, Saori Takahashi, … Ichiro Hiratani

- Article
- Open Access
- Published: 15 March 2023

Whole-genome doubling drives oncogenic loss of chromatin segregation

- Ruxandra A. Lambuta,
- Luca Nanni,
- Yuanlong Liu,
- Juan Diaz-Miyar,
- Arvind Iyer,
- Daniele Tavernari,
- Natalya Katanayeva,
- Giovanni Ciriello &
- …
- Elisa Oricchio

Show authors

Nature **volume 615**, pages 925–933 (2023)

- 16k Accesses
- 272 Altmetric
- Metrics details

Abstract

Whole-genome doubling (WGD) is a recurrent event in human cancers and it promotes chromosomal instability and acquisition of aneuploidies[1,2,3,4,5,6,7,8]. However, the three-dimensional organization of chromatin in WGD cells and its contribution to oncogenic phenotypes are currently unknown. Here we show that in p53-deficient cells, WGD induces loss of chromatin segregation (LCS). This event is characterized by reduced segregation between short and long chromosomes, A and B

subcompartments and adjacent chromatin domains. LCS is driven by the downregulation of CTCF and H3K9me3 in cells that bypassed activation of the tetraploid checkpoint. Longitudinal analyses revealed that LCS primes genomic regions for subcompartment repositioning in WGD cells. This results in chromatin and epigenetic changes associated with oncogene activation in tumours ensuing from WGD cells. Notably, subcompartment repositioning events were largely independent of chromosomal alterations, which indicates that these were complementary mechanisms contributing to tumour development and progression. Overall, LCS initiates chromatin conformation changes that ultimately result in oncogenic epigenetic and transcriptional modifications, which suggests that chromatin evolution is a hallmark of WGD-driven cancer.

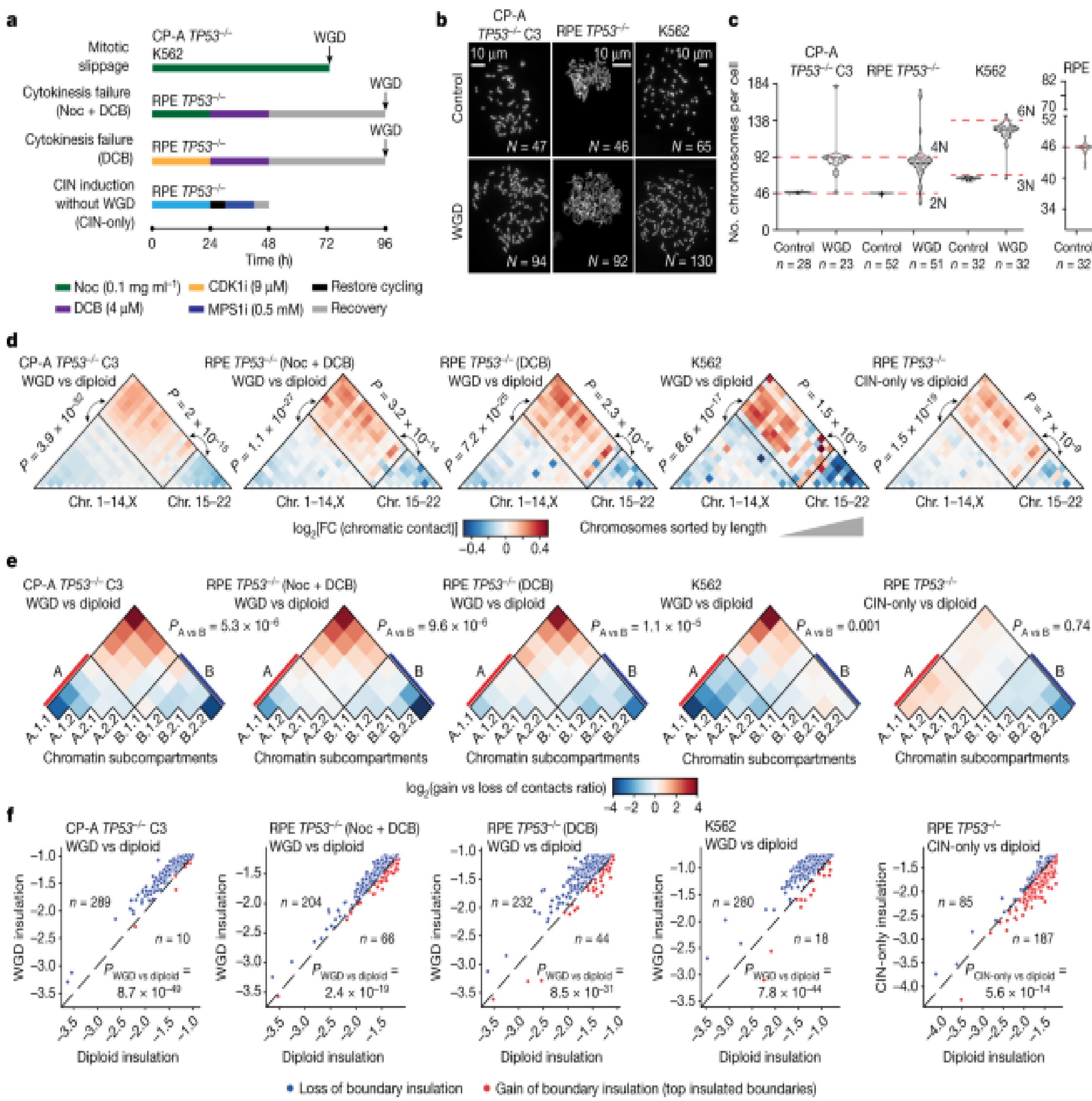
Main

WGD is defined by the duplication of the entire set of chromosomes within a cell. It has been observed in early and pre-malignant lesions of various tissues[2,9,10], and it is estimated to occur in approximately 30% of human cancers[3]. WGD favours the acquisition of chromosomal alterations[5,6,7,8] in permissive genetic backgrounds, such as in p53- or Rb-deficient cells[3,4], which may promote tumorigenesis[1,5]. However, tetraploidization in single nuclei is equally likely to induce alterations in the three-dimensional (3D) structure and epigenetic features of the chromatin. During interphase, chromatin is organized in a multilayer 3D architecture of compartments, chromatin domains, and loops[11,12,13,14,15,16], and is closely associated with chromatin activity and cell states[17]. Alterations of the chromatin structure have been reported in many tumour types and are due to altered CTCF or cohesin binding[18,19], chromosome structural variants[20,21,22] or aberrant histone modifications[22,23,24,25]. Here we investigate how chromatin is organized in cells that undergo WGD. Moreover, we study which features of the chromatin structure are affected by WGD and whether changes in chromatin organization emerge and affect cell phenotypes after WGD. Finally, we examine whether these changes correlate with genetic and epigenetic alterations in WGD-driven tumours.

WGD results in LCS

To understand the impact of WGD on chromatin organization and tumour development, we used three distinct cellular models: (1) the non-transformed diploid cell line hTERT-RPE1 (hereafter referred to as RPE); (2) CP-A cells derived from a patient with Barrett's oesophagus, a pre-cancerous condition that predisposes to oesophageal adenocarcinoma development through WGD[9,26]; and (3) the leukaemic near triploid K562 cell line. To mimic the permissive genetic background observed in human tumours[3,26], we used p53-deficient CP-A (Extended Data Fig. 1a) and RPE cells (previously termed RPE$^{TP53-/-}$ cells)[27]. K562 cells already harbour a loss-of-function mutation in the *TP53* gene[28]. WGD cells were obtained through mitotic slippage in two independent CP-A *TP53*$^{-/-}$ clones (clone 3 and clone 19) and K562 cells, and through cytokinesis failure using two distinct protocols in RPE *TP53*$^{-/-}$ cells (Fig. 1a). To control for chromatin conformation changes associated with chromosomal instability (CIN) but not WGD, we induced CIN in RPE *TP53*$^{-/-}$ cells using a MPS1 inhibitor (Fig. 1a). Cell cycle and karyotype analyses confirmed that the number of chromosomes doubled after treatment in most cells (Fig. 1b,c and Extended Data Fig. 1b–g) and that the nuclear size increased (Extended Data Fig. 1h).

Fig. 1: WGD induces LCS.



a, Schematic representation of the WGD and CIN induction experimental approaches in the indicated cell lines. DCB, dihydrocytochalasin B; Noc, nocodazole. **b**, Representative images of metaphase spreads for CP-A *TP53*$^{-/-}$ clone 3 (C3), RPE *TP53*$^{-/-}$ and K562 cell lines. **c**, Quantification of chromosomes per cell for CP-A C3 *TP53*$^{-/-}$, RPE *TP53*$^{-/-}$ and K562 cell lines. The number of cells is indicated. For the violin plots, dashed line is the median, dotted lines are quartiles. **d**, Heatmap of the ratios of interchromosomal contact enrichments (observed versus expected) between WGD and control samples for CP-A *TP53*$^{-/-}$ cells, RPE *TP53*$^{-/-}$ cells and K562 cells and between RPE *TP53*$^{-/-}$ CIN-only cells and control cells. Chromosomes were sorted by length. **e**, Heatmap of ratios of genomic bins belonging to the indicated subcompartments that gain or lose contacts in the

indicated conditions. **f**, Boundary insulation scores in control cells and WGD or CIN-only cells in the indicated cell lines for the shared top insulating boundaries. For **f–h**, *P* values were calculated using two-tailed Wilcoxon test.

Source data

Conversely, treatment with the MPS1 inhibitor did not change the ploidy of the RPE $TP53^{-/-}$ cell population (hereafter, CIN-only RPE $TP53^{-/-}$), but the cells exhibited a variable number of chromosomes (Fig. 1c). Hence, we analysed chromatin organization before and after WGD induction through high-throughput chromatin conformation capture (Hi-C) analysis in all models (Supplementary Fig. 1). Despite the doubling of the genome, chromatin organization was highly similar between diploid and tetraploid cells (Extended Data Fig. 2a–e). However, the ratios of the observed number of contacts compared with the expected number of contacts at each locus indicated that the enrichment and depletion of contacts were lower in WGD cells than in diploid cells. By contrast, these ratios remained similar in CIN-only and diploid cells (Extended Data Fig. 2f,g). In particular, the number of contacts within a domain or compartment decreased, whereas the number of contacts between different domains and compartments increased. To further investigate the changes in chromatin contact distribution in WGD cells, we assessed the following parameters: (1) changes in contacts between the clusters of long and short chromosomes[11]; (2) contact enrichment within A and B subcompartments, which we inferred using the Calder algorithm[29]; and (3) contact insulation at topologically associating domain (TAD) boundaries[30]. In all WGD-induction models and independent replicates, compared with control cells, WGD cells consistently exhibited the following characteristics: (1) a significantly increased proportion of contacts between long chromosomes (1–14 and X) and short chromosomes (15–22) (Fig. 1d and Extended Data Fig. 3a,b); (2) a significantly increased proportion of contacts between A and B compartments, especially between the most distant A.1.1 and B.2.2 (Fig. 1e and Extended Data Fig. 3c); and (3) decreased boundary insulation (Fig. 1f and Extended Data Fig. 3d). These effects were only moderately detectable or absent in CIN-only cells (Fig. 1d–f) and did not depend on the resolution of the Hi-C experiment, coverage

per haploid copy or ratios of contacts between homologous copies of the same chromosomes (Extended Data Fig. 3e,f).

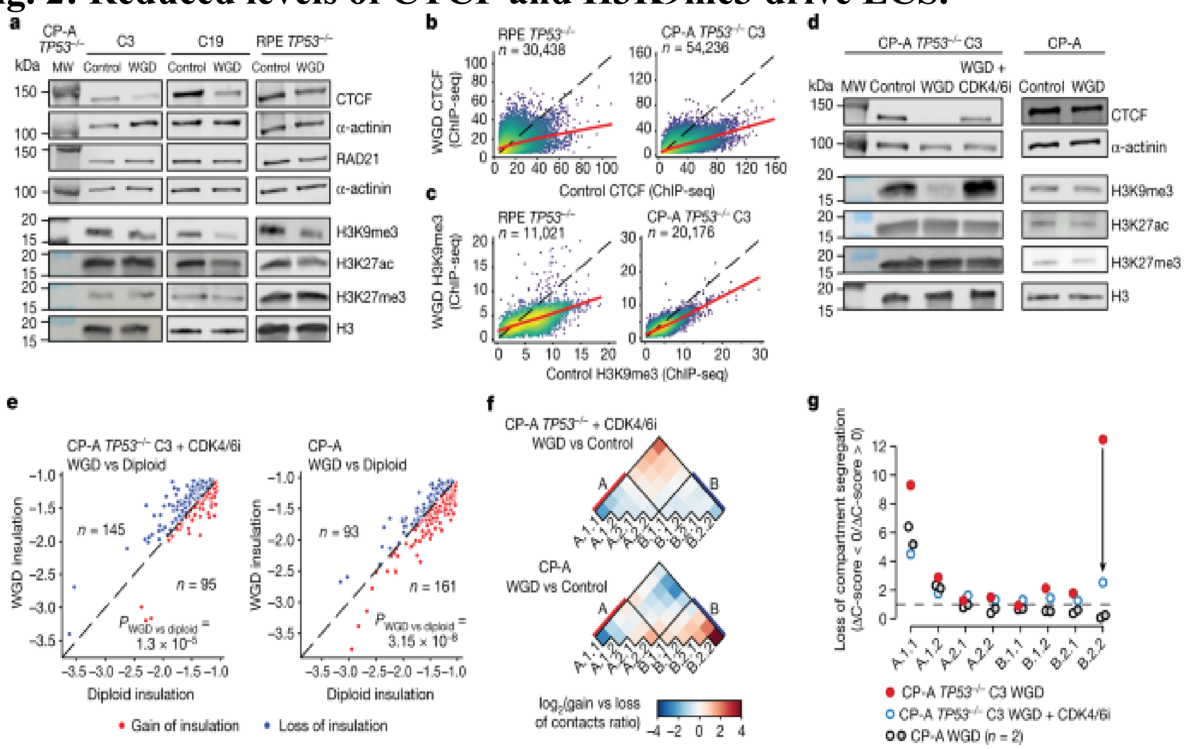
Gene expression analysis comparing RPE *TP53*$^{-/-}$ WGD cells and control cells revealed an overall upregulation of transcription in WGD cells (Extended Data Fig. 4a). Significantly upregulated genes ($n = 1,268$, $\log_2$(fold change (FC)) > 1, adjusted $P < 0.01$) were enriched in the interferon signalling pathway (Extended Data Fig. 4b), which is consistent with responses to abnormal mitotic segregation and stress[31]. Conversely, significantly downregulated genes ($n = 619$, $\log_2$(FC) < −1, adjusted $P < 0.01$) were enriched in the DNA replication, DNA repair and cell cycle pathways (Extended Data Fig. 4c), which is consistent with downregulation of DNA replication proteins in WGD cells[7]. Changes in expression were not associated with changes in compartment segregation and only moderately with boundary loss of insulation (Extended Data Fig. 4d,e), which indicated that these changes mostly reflected an acute cell response to WGD. In summary, WGD cells, but not CIN-only cells, exhibit LCS manifested in an increased proportion of contacts between long and short chromosomes, distinct chromatin subcompartments, and TADs.

CTCF and H3K9me3 deficiency determines LCS

Increased contact frequency among long and short chromosomes could be associated with the doubled number of homologous chromosomes. Therefore, we investigated causes of boundary insulation loss and loss of compartment segregation. CTCF and cohesin are crucial proteins for maintaining insulation at TAD boundaries[32,33], whereas enrichment of specific histone marks is associated with chromatin compartmentalization[14,29]. CP-A *TP53*$^{-/-}$ cells and RPE *TP53*$^{-/-}$ cells that underwent WGD exhibited an approximately 50% reduction in CTCF and H3K9me3 compared with control cells. WGD cells also had a modest decrease in H3K27ac, but no consistent changes in H3K27me3 or the cohesin complex component RAD21 (Fig. 2a and Extended Data Fig. 4f). *CTCF* mRNA abundance was also lower in WGD cells than in diploid control cells ($\log_2$(FC) = −0.6, adjusted $P = 8.3 \times 10^{-6}$) (Supplementary Table 1). Chromatin immunoprecipitation with high-throughput and

sequencing (ChIP–seq) analysis of CTCF showed that WGD cells and diploid cells shared the majority of CTCF peaks (Extended Data Fig. 4g); however, these peaks typically exhibited lower signal (input-normalized number of reads) in WGD cells than in diploid cells (Fig. 2b and Extended Data Fig. 4h). TAD boundaries that lost insulation showed lower CTCF abundance and fewer numbers of CTCF peaks compared with boundaries that retained or even gain insulation in WGD cells (Extended Data Fig. 4i). This result suggests that reduced CTCF protein levels lead to a stochastic loss of CTCF binding, which in turn results in loss of insulation at boundaries with few CTCF binding sites. In parallel, ChIP–seq analysis of H3K9me3 levels confirmed an overall reduction in WGD cells, particularly at regions that originally exhibited high H3K9me3 levels (Fig. 2c) and in the B.2.2 subcompartment, which is usually enriched for this histone mark (Extended Data Fig. 4j).

Fig. 2: Reduced levels of CTCF and H3K9me3 drive LCS.



a, Representative images of immunoblots of indicated proteins in diploid control and WGD CP-A *TP53*$^{-/-}$ (clone 3 and clone 19 (C19)) and RPE *TP53*$^{-/-}$ cells. MW, molecular weight **b**,**c**, ChIP–eq signal (FC over input) of CTCF (**b**) and H3K9me3 (**c**) shared peaks between control and

WGD conditions in CP-A *TP53*$^{-/-}$ clone 3 and RPE *TP53*$^{-/-}$ cells. Dots are coloured by point density in $\log_{10}$ scale, and the regression line is in red. **d**, Representative immunoblots of indicated proteins in diploid control and WGD CP-A cells and in WGD CP-A *TP53*$^{-/-}$ cells treated with a CDK4/6i. **e**, Boundary insulation scores in control and WGD cells for the shared top insulating boundaries between the indicated conditions. *P* values were calculated using two-tailed Wilcoxon test. **f**, Heat map of ratios of genomic bins belonging to the indicated subcompartments that gain versus lose contacts in the indicated conditions. **g**, Loss of compartment segregation score relative to the control condition in each subcompartment domain in the indicated cell lines.

Source data

In our models, the lack of p53 creates a permissive genetic background that allows WGD cells to bypass the tetraploid checkpoint, tolerate DNA damage and continue to grow[1,4,34]. Thus, we tested whether the lack of checkpoints and uncontrolled proliferation of *TP53*$^{-/-}$ cells contribute to the inability of WGD cells to increase CTCF and H3K9me3 levels. First, we induced WGD in *TP53* wild-type cells that activate the tetraploid checkpoint, which stalls cells in the G1 cell cycle phase. Second, we induced WGD in CP-A *TP53*$^{-/-}$ cells treated with an inhibitor of CDK4 and CDK6 (CDK4/6i), palbociclib, which leads to a prolonged G1 phase (Extended Data Fig. 4k). We successfully induced WGD in *TP53* wild-type CP-A cells (Extended Data Fig. 4l–n), whereas most *TP53* wild-type RPE cells remained binucleated after treatment with nocodazole and dihydrocytochalasin B and could not be used for further analyses (Extended Data Fig. 4o). In *TP53* wild-type cells, normalized CTCF and H3K9me3 levels were comparable between WGD and diploid cells, and treatment with palbociclib was sufficient to rescue CTCF and H3K9me3 levels in *TP53*$^{-/-}$ WGD cells (Fig. 2d and Extended Data Fig. 4p). After rescuing CTCF and H3K9me3 levels, loss of insulation at TAD boundaries and loss of compartment segregation was strongly reduced or completely absent (Fig. 2e,f) compared with what was observed in CP-A *TP53*$^{-/-}$ WGD cells (Fig. 1e,f), in particular within the B.2.2 subcompartment (Fig. 2g). Loss of segregation between long and short chromosomes remained detectable in *TP53* wild-type cells and in WGD cells treated with palbociclib (Extended Data Fig. 4q). This result indicates

that this effect is independent of p53, CTCF and H3K9me3 status, and is probably due to the doubled number of chromosomes.

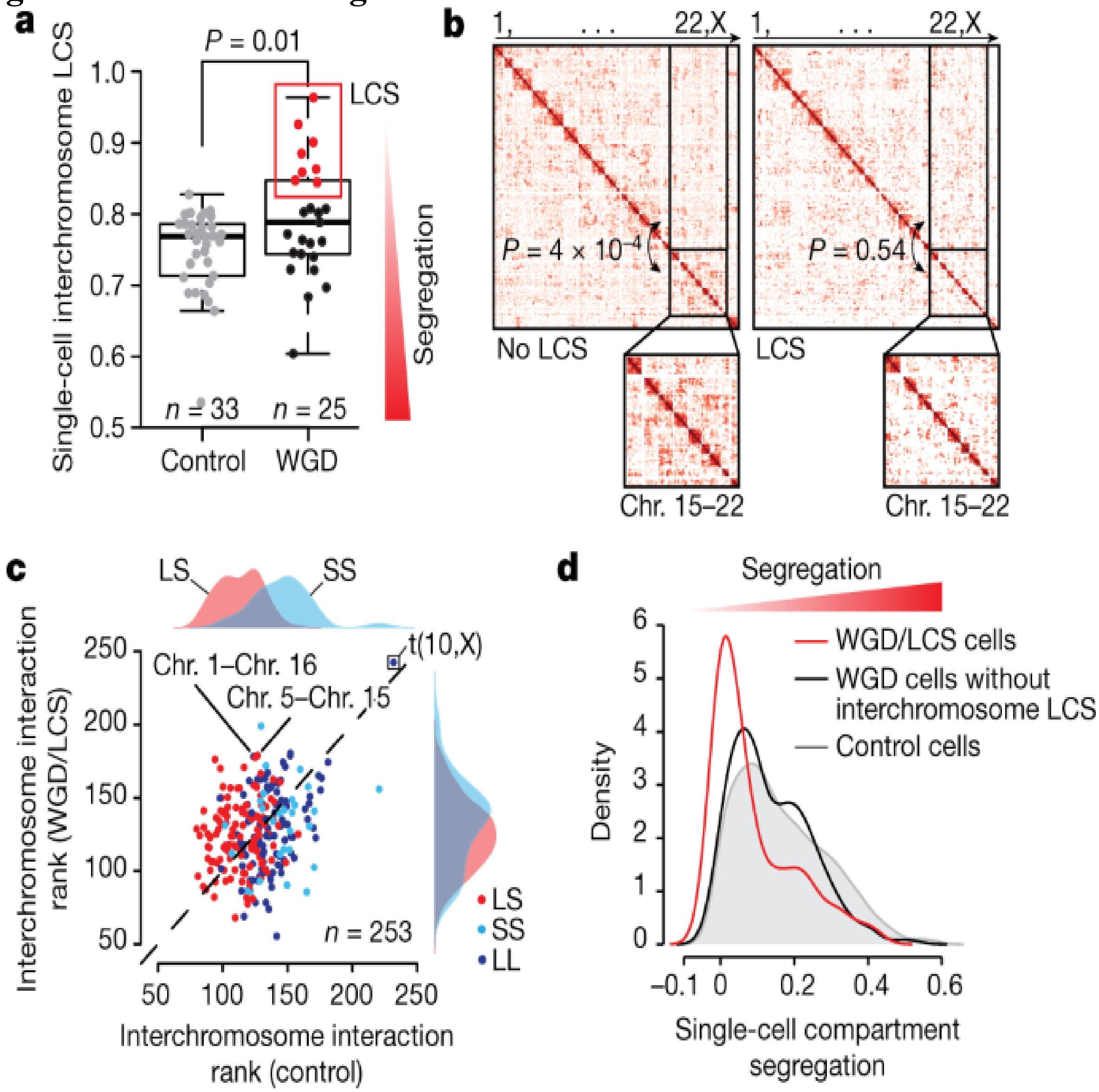
Although *TP53* loss was required to induce LCS after WGD, LCS was not detectable when comparing diploid *TP53*$^{-/-}$ cells with diploid *TP53* wild-type cells (Extended Data Fig. 5a–c). CTCF protein expression was also retained (Extended Data Fig. 5d), which is not a direct target of *TP53* (Extended Data Fig. 5e). These data indicate that activation of the p53-dependent tetraploid checkpoint is important to increase protein production and to maintain chromatin conformation and epigenetic status in WGD cells.

LCS is detectable in WGD single cells

Next we asked whether loss of segregation among chromosomes, compartments, and chromatin domains could also be detected in single cells. We performed single-cell Hi-C (scHi-C) in RPE *TP53*$^{-/-}$ diploid cells and WGD cells by isolating individual nuclei from the two cell populations (Supplementary Fig. 2). scHi-C libraries were prepared from 73 individual nuclei, and, after sequencing, we retained 33 control cells and 25 WGD cells (Supplementary Table 2; mean number of contacts per cell = 565,324). Aggregating scHi-C profiles (pseudo-bulk) reproduced the enrichment patterns observed in bulk RPE *TP53*$^{-/-}$ Hi-C data (Extended Data Fig. 6a). After comparing the number of contacts between long and short chromosomes and among long chromosomes and short chromosomes, a subpopulation of cells exclusively detectable in the WGD group exhibited an increased proportion of interactions among long and short chromosomes (Fig. 3a). Short–short chromosome contacts were significantly enriched compared with long–short chromosome contacts in cells that did not exhibit LCS, but this difference was no longer detectable in LCS-exhibiting WGD (LCS-WGD) cells (Fig. 3b, Supplementary Fig. 2 and Supplementary Table 2). By ranking chromosome pairs on the basis of their total number of interchromosomal contacts, chromosome 10 and the X chromosome scored at the top in both WGD cells and control cells, consistent with a t(10,X) translocation reported in RPE cells[35] (Fig. 3c). Notably, long–short chromosome pairs obtained lower ranks than short–short chromosome pairs in control cells, but not in LCS-WGD cells. For these cells, the top scoring

chromosome pairs included pairs such as chromosome 1–chromosome 16 and chromosome 5–chromosome 15 (Fig. 3c).

Fig. 3: LCS in WGD single cells.



a, Interchromosome LCS score calculated for each cell. Black dots indicate WGD cells not exhibiting LCS, red dots indicate WGD cells exhibiting LCS. For the boxplots, the central line is the median, the bounding box corresponds to the 25–75th percentiles and the whiskers extend up to 1.5-times the interquartile range. *P* value was calculated using two-tailed Wilcoxon test. **b**, Representative Hi-C maps at 10 Mb resolution of WGD RPE *TP53*$^{-/-}$ cells with (right) and without LCS (left). Short–short (SS) and

long–short (LS) chromosome contacts were compared using two-tailed Wilcoxon test. **c**, Average interchromosomal interaction rank across single cells for each pair of chromosomes in control and WGD-LCS cells. The chromosome pair (10,X) is highlighted. **d**, Single-cell compartment segregation score distribution.

Source data

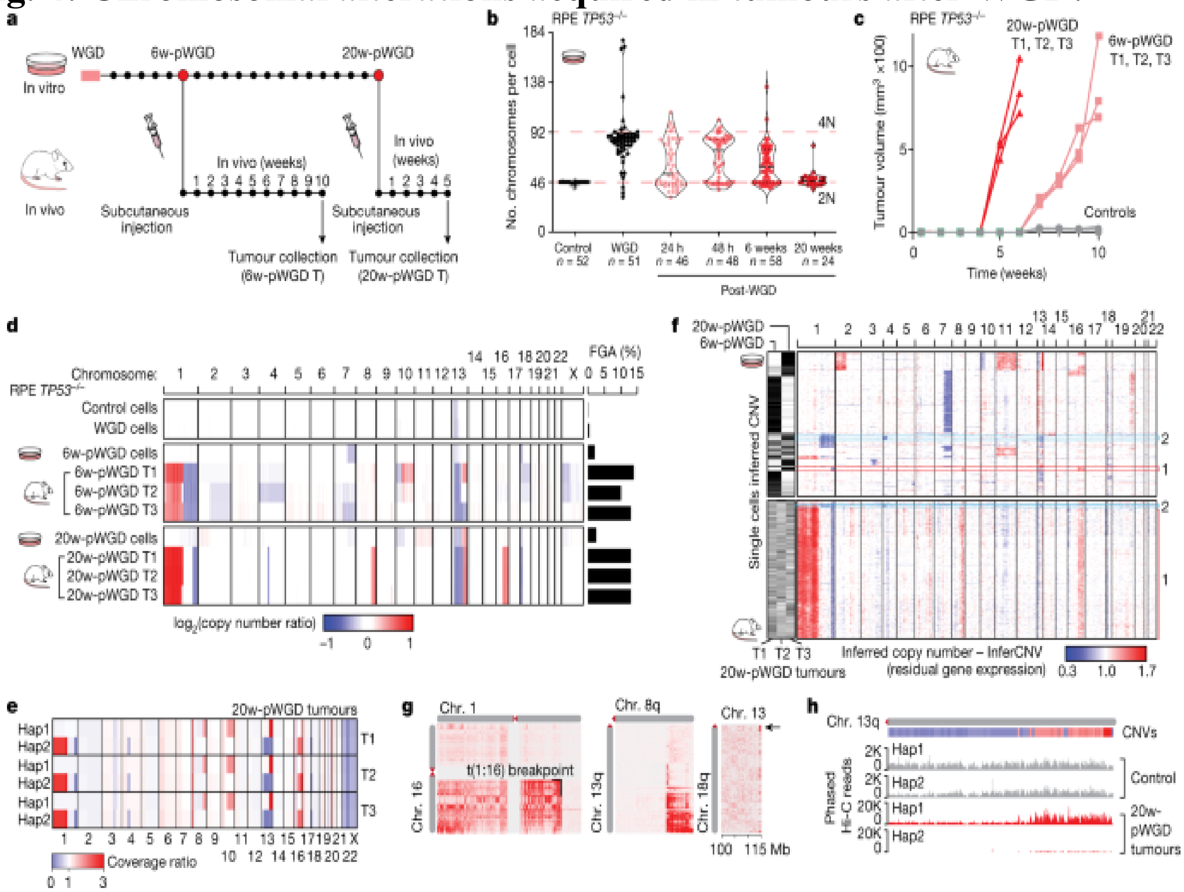
Next we inferred A and B compartments in single cells to assess compartment segregation. LCS-WGD cells also exhibited significantly reduced compartment segregation (Fig. 3d), which indicated that the LCS features observed at the population level are intrinsically present within this group of single cells. The sparsity of the scHi-C data did not enable the assessment of boundary insulation. Last, copy number variants (CNVs) inferred from scHi-C coverage showed that LCS in WGD cells did not associate with the number of CNVs or the fraction of the genome altered (FGA) (Extended Data Fig. 6b,c). Hence, LCS could be detected in single cells and did not depend on CNV acquisition.

Genomic evolution of WGD cells

Following WGD, both RPE *TP53*$^{-/-}$ cells and CP-A *TP53*$^{-/-}$ cells showed a transition to a heterogenous and aneuploid karyotype within 48 h (Extended Data Fig. 7a–c), which is consistent with WGD and loss of p53 favouring aneuploidy and CIN[1,9,36]. As early as 24 h after WGD (post-WGD), we detected CIN characteristics, such as chromosome breakages and telomere fusions in CP-A *TP53*$^{-/-}$ cells (Extended Data Fig. 7d), and multipolar spindles and bipolar division with clustered centrosomes in RPE *TP53*$^{-/-}$ cells (Extended Data Fig. 7e). To elucidate the evolution of CIN and chromatin 3D structures in post-WGD cells, we analysed genomic and chromatin conformation changes in RPE *TP53*$^{-/-}$ cell populations at different time points in vitro (up to 20 weeks) and in vivo (Fig. 4a). At 6 weeks post-WGD, RPE *TP53*$^{-/-}$ cells in vitro exhibited heterogeneous ploidy, whereas the population became nearly diploid at 20 weeks post-WGD (Fig. 4b). At these two time points, cells were subcutaneously injected into immunocompromised mice. All animals engrafted with RPE *TP53*$^{-/-}$

cells at 6 weeks post-WGD ($n = 12$) or 20 weeks post-WGD ($n = 6$) developed tumours within 2.5 and 1.5 months, respectively (Fig. 4c and Extended Data Fig. 8a). By contrast, RPE *TP53*$^{-/-}$ diploid cells did not induce tumorigenesis (Fig. 4c), which indicated that the oncogenic capacity of these cells was acquired after WGD.

Fig. 4: Chromosomal alterations acquired in tumours after WGD.



a, Timeline of in vitro and in vivo experiments for long-term post-WGD cells. **b**, Chromosome per cell counts in RPE *TP53*$^{-/-}$ cells of control, WGD, 6-weeks post-WGD (6w-pWGD) and 20-weeks post-WGD (20w-pWGD) populations. The number of cells considered for each condition is indicated. **c**, Tumour volumes ($mm^3 \times 100$) from the time of subcutaneous injection in NOD SCID gamma (NSG) mice of RPE *TP53*$^{-/-}$ control ($n = 3$), 6-weeks post-WGD ($n = 3$) and 20-weeks post-WGD ($n = 3$) cells. **d**, Copy number alterations determined by WGS data in RPE *TP53*$^{-/-}$ control, WGD and post-WGD samples. Bar plots show the FGA for each sample **e**, Haplotype resolved copy number profile of each of the three 20-weeks post-WGD

tumours. **f**, Copy number profile for each single cell inferred from scRNA-seq data in RPE *TP53*$^{-/-}$ 6-weeks post-WGD, 20-weeks post-WGD and 20-weeks post-WGD tumours. The sample of origin of each cell is indicated on the left. Clonal populations detected in tumours and in vitro samples are highlighted (clones 1 and 2). **g**, Interchromosomal Hi-C contact maps exhibiting contact patterns consistent with chromosomal translocations in the RPE *TP53*$^{-/-}$ 20-weeks post-WGD derived tumour 2 (T2). Three chromosomal translocations are highlighted: t(1:16), t(8:13) and t(13:18). **h**, Distribution of phased Hi-C reads between haplotype 1 (Hap1) and haplotype 2 (Hap2) in RPE *TP53*$^{-/-}$ control and 20-weeks post-WGD derived tumour 2 samples for chromosome 13q. The corresponding copy number status for tumour 2 is shown on the top (red, copy number gains; blue, copy number losses).

Source data

We next performed whole-genome sequencing (WGS) analyses of in vitro and in vivo post-WGD samples. The data showed that the number of acquired mutations in 6-weeks post-WGD RPE *TP53*$^{-/-}$ cells was about 1.8-times higher than in control cells kept in culture for the same amount of time, and post-WGD mutations had lower variant allele frequencies (Extended Data Fig. 8b). Across all samples, we detected a heterozygous clonal *NRAS* Q61R mutation (variant allele frequency > 0.4), which is a known oncogenic variant[37]. Nevertheless, this mutation was already present in RPE *TP53*$^{-/-}$ cells before WGD, and it was not sufficient to induce tumorigenesis in mice (Fig. 4c). Conversely, mutations that were acquired post-WGD did not include known oncogenic variants (Supplementary Table 3).

RPE *TP53*$^{-/-}$ diploid cells and WGD cells exhibited a nearly unaltered genome (FGA < 1%), except for a shallow loss on chromosome 13p (Fig. 4d and Supplementary Table 3). In vitro samples at 6-weeks and 20-weeks post-WGD exhibited evidence of acquired CNVs (FGA = 2% and 2.5%, respectively), although a higher number of CNVs became evident only in the in vivo tumour samples generated from either 6-week or 20-week post-WGD cells (Fig. 4d; mean FGA = 12% and 13%, respectively) (Supplementary Table 4). Shared CNV breakpoints and altered haplotypes

indicated that tumours derived from 20-week post-WGD RPE *TP53*$^{-/-}$ cells originated from the selection and expansion of the same clone in vivo (Fig. 4d,e and Extended Data Fig. 8c). CNV acquisition was observed in nine additional tumours originated from three independent WGD experiments (Extended Data Fig. 8d). Notably, tumours derived from independent experiments sometimes acquired similar CNVs, which indicated the occurrence of convergent evolution, as recently observed in animal models after a transient induction of CIN[38].

The relatively low number of CNVs detected in the in vitro samples could be explained by subclonal heterogeneity. To test this hypothesis, we analysed all samples by single-cell RNA-sequencing (scRNA-seq) and inferred the copy number status from the read sequencing depth using the algorithm InferCNV[39]. InferCNV analysis revealed that 6-week and 20-week post-WGD in vitro samples exhibited highly heterogenous copy number changes and clustered in distinct subclones, which were present in different proportions in the two samples (Fig. 4f and Extended Data Fig. 8e). By contrast, tumour samples derived from 20-week post-WGD cells were largely composed of a single clone (Fig. 4f), which exhibited CNVs consistent with those detected by WGS analyses and could already be detected in vitro, along with a less prevalent one (Fig. 4f, clones 1 and 2 on the right). Beyond CNVs, analysis of WGS and Hi-C data from these tumour samples revealed a new chromosomal translocation between chromosomes 1 and 16 (Fig. 4g), which were among the chromosome pairs that had the most increased contact frequency in LCS-WGD cells (Fig. 3c). Moreover, two translocations involving chromosome 13, one with chromosome 8 and one with chromosome 18, the latter involving the telomeric region of chromosome 13, were also observed (Fig. 4g). Notably, we could not find evidence of these translocations in control cells or WGD cells, which indicated that these events were acquired after WGD. Loss of the telomeric end in chromosome 13 was accompanied by complex chromosomal rearrangements on the second part of the q arm, and involved alternating high copy number gains (up to five copy gains) and copy number losses (Extended Data Fig. 8f). This rearrangement pattern is characteristic of multiple breakage–fusion–bridge cycles[40,41]. Moreover, all these chromosomal rearrangements occurred in only one of the two haplotypes (Hap1), whereas the other was lost (Hap2) (Fig. 4h).

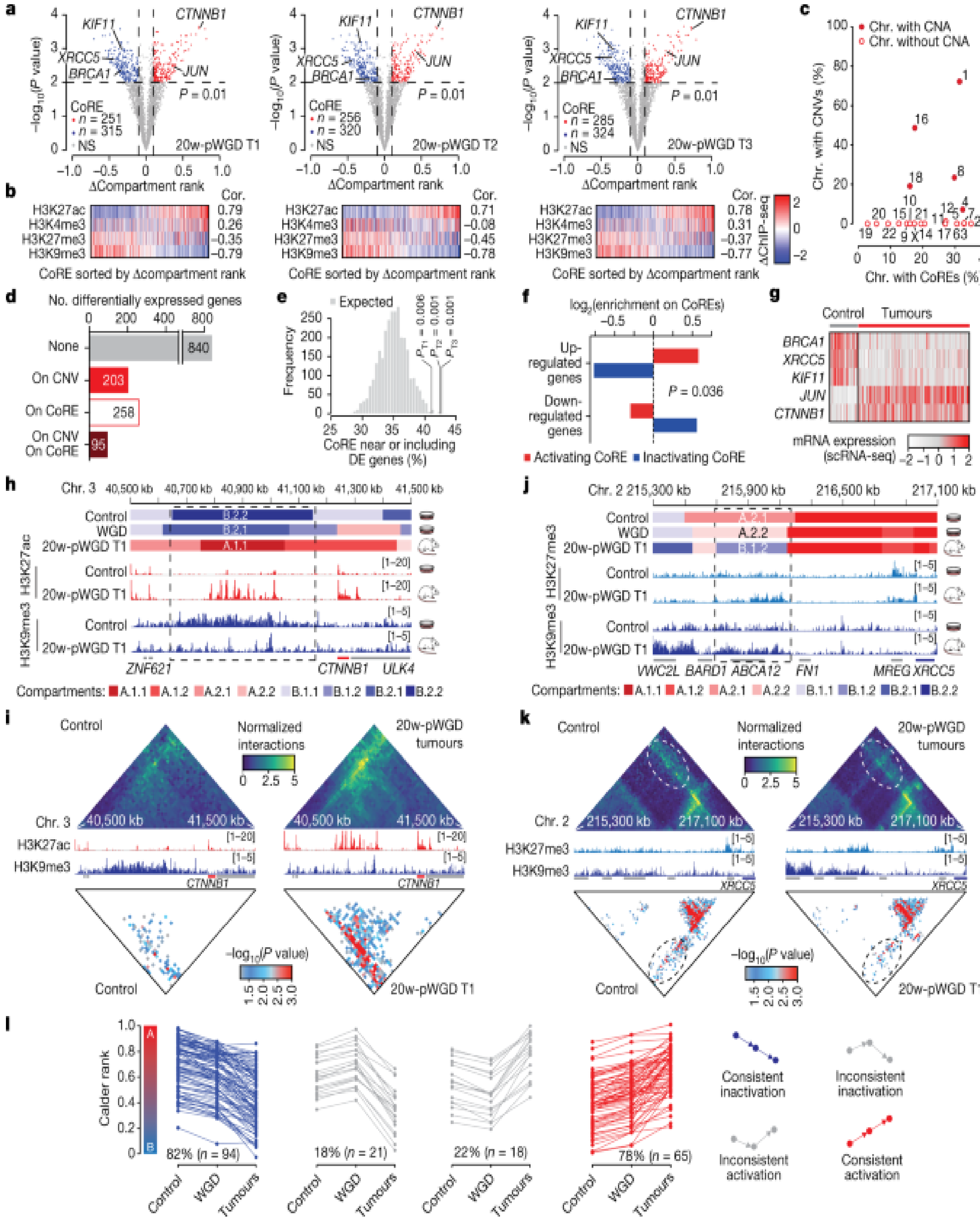
Copy number losses or gains determined by WGS analyses were associated with reduced and increased gene expression, respectively, as estimated by scRNA-seq (Extended Data Fig. 8g and Supplementary Table 5). These losses and gains accounted for around 20% of differentially expressed genes (adjusted $P < 0.001$, absolute $\log_2(FC) > 0.3$). These changes comprised upregulation of inducers of cell proliferation and migration such as *CDC42*, *NRAS* and *JUN* (chromosome 1p)[42,43], and downregulation of *CENPF* (chromosome 1q), which is associated with mitotic errors[44]. *NRAS* copy number gain was accompanied by an increase in Q61R variant allele frequency ($VAF_{tumour1} = 0.9$, $VAF_{tumour2} = 0.75$, $VAF_{tumour3} = 0.74$), which suggested that the mutated allele was in the amplified haplotype. *JUN* overexpression was concomitant with an upregulation of components of the AP-1 transcription factor complex (*JUND*, *JUNB*, *FOS* and *FOSB*) and its downstream targets (Extended Data Fig. 8h). In summary, tumours originated from RPE *TP53*$^{-/-}$ cells that underwent WGD exhibited hallmarks of WGD-driven human tumours, such as increased CIN and complex rearrangements potentially associated with oncogene activation.

Chromatin evolution of WGD cells

Next we investigated the long-term effects of WGD on chromatin 3D organization and its functional consequences. Hi-C analyses showed that tumours generated from 20-week post-WGD cells partially retained LCS features (Extended Data Fig. 9 and Supplementary Fig. 3), although these could be confounded by the high number of aneuploidies and changes in chromatin organization. Indeed, compared with RPE *TP53*$^{-/-}$ control cells, tumour samples exhibited greater differences than WGD cells in both compartment domain ranks and subcompartment assignments inferred using Calder[29] (Extended Data Fig. 10a,b). By developing a new algorithmic approach, we searched for regions that significantly changed subcompartment (Extended Data Fig. 10c), termed compartment repositioning events (CoREs). In total, we found 487 (tumour 1), 481 (tumour 2) and 478 (tumour 3) significant CoREs, which indicated changes towards either a more active (activating CoRE) or a more inactive (inactivating CoRE) subcompartment (Fig. 5a, Extended Data Fig. 10d,e and Supplementary Table 6). Genome-wide subcompartment changes and

CoREs correlated with changes in histone mark intensities (Fig. 5b and Extended Data Fig. 10f,g), particularly H3K9me3 and H3K27ac, which suggested that they could underlie changes in regulatory interactions[22]. CoREs covered 17–18% of the genome and were found in similar proportions in chromosomes affected or unaffected by CNVs (Fig. 5c and Extended Data Fig. 10h). CoREs detected using our algorithm were largely recapitulated using an independent strategy (Extended Data Fig. 10i–l). Differentially expressed genes between tumours and RPE *TP53*$^{-/-}$ control cells were observed in similar numbers within a CNV or within a CoRE (Fig. 5d). CoREs were more likely to include or be near (<1 Mb) a differentially expressed gene than randomly selected genomic regions of the same size (Fig. 5e). Moreover, upregulated and downregulated genes were enriched in CoREs that changed towards a more active or inactive compartment, respectively (Fig. 5f). For example, we found activating CoREs in correspondence with upregulated oncogenes such as *JUN*, which was also amplified, and β-catenin (encoded by *CTNNB1*)[45], which was among the most significant CoREs in all three tumour samples (Fig. 5a,g). By contrast, inactivating CoREs comprised downregulated tumour suppressors and DNA repair genes such as *BRCA1* and *XRCC5* (refs. [46,47]), and the kinesin family member *KIF11*, the loss of which is associated with CIN[48] (Fig. 5a,g). The CoRE associated with *CTNNB1* was upstream of the gene and corresponded to a change from the most inactive subcompartment (B.2.2) in RPE *TP53*$^{-/-}$ diploid cells to the most active subcompartment (A.1.1) in all three tumour samples (Fig. 5h, top, and Extended Data Fig. 11a). Within this CoRE in the tumour samples, we detected the formation of multiple H3K27ac peaks and a reduction in H3K9me3, but minor or no changes in CTCF and other histone marks (Fig. 5h and Extended Data Fig. 11a). Accumulation of H3K27ac indicated the formation of a new large enhancer, and it was associated with increased contact frequencies and significant interactions between the *CTNNB1* promoter and the enhancer region (Fig. 5i). A similar formation of H3K27ac peaks and enhancer–promoter interactions were found in a CoRE downstream of *JUN* (Extended Data Fig. 11b,c), which indicated a synergistic activation of the oncogene mediated by whole-arm chromosome gain (chromosome 1p; Fig. 4d), subcompartment repositioning, and histone acetylation changes.

Fig. 5: Comparison of subcompartment repositioning in tumours after WGD.



a, Volcano plots of segmented genomic regions between each tumour and control cells. Selected CoREs are labelled on the basis of genes overlapping or in the proximity (±1 Mb) of the region. *P* values calculated using

DiffComp. NS, not significant. **b**, Differential ChIP–seq signal in CoRE regions between control samples and each 20-weeks post-WGD tumour. **c**, Correlation between percentage of chromosomes affected by CoREs and CNVs for each chromosome in RPE *TP53*$^{-/-}$ 20-weeks post-WGD tumours. **d**, The number of differentially expressed genes in regions unaffected (None) or affected by CNVs, CoREs or both. **e**, Expected and observed percentage of CoREs near to (±1 Mb) or overlapping with differentially expressed genes in the RPE *TP53*$^{-/-}$ 20-weeks post-WGD tumours and control samples. **f**, Enrichment of differentially expressed genes in 20-weeks post-WGD tumours versus control in activating or inactivating CoREs. **g**, Normalized expression levels in single cells of selected differentially expressed genes between RPE *TP53*$^{-/-}$ control samples and 20-weeks post-WGD tumours. **h**,**j**, Detailed characterization of the compartment and histone modification changes in the regions of chromosome 3 (**h**) and chromosome 2 (**j**) in RPE *TP53*$^{-/-}$ control and tumour 1. Top, subcompartment assignments inferred by Calder. Bottom, histone mark intensities. **i,k**, Distance-normalized interaction maps at 25 kb resolution in the regions of chromosome 3 (**i**) and chromosome 21 (**j**) in RPE *TP53*$^{-/-}$ control and tumour samples (top). Histone mark intensities for the corresponding sample (middle), significant interactions RPE *TP53*$^{-/-}$ control and tumour 1 samples (bottom). *P* values calculated using HiC-DC **l**, Compartment rank in control, WGD and tumours for each activating and inactivating CoRE region. Lines connect compartment ranks belonging to the same CoRE region. For **e** and **f**, *P* values were derived by data permutation ($n = 1{,}000$).

Source data

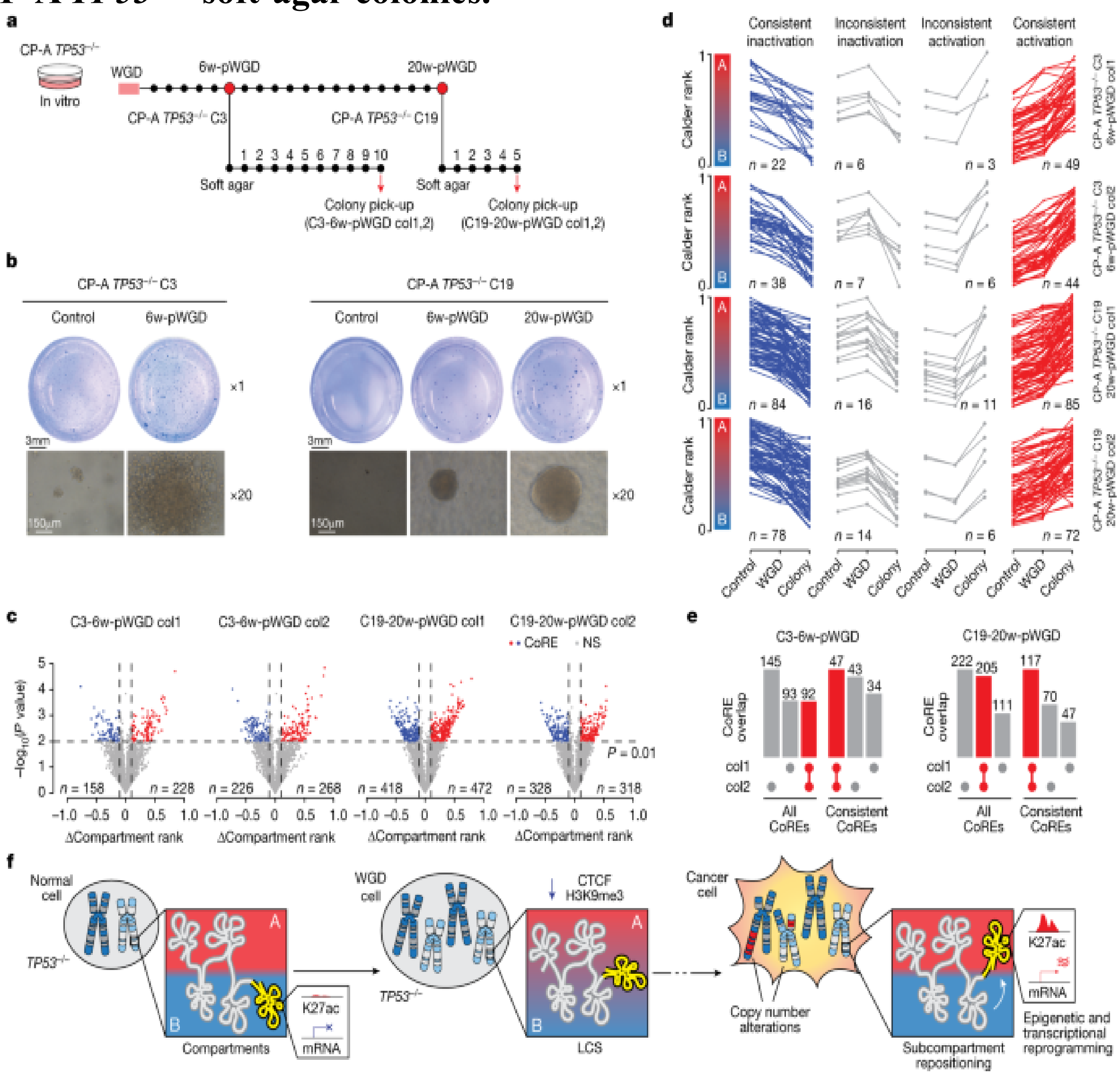
Next we examined subcompartment repositioning involving the tumour suppressors *XRCC5* (Fig. 5j and Extended Data Fig. 11d) and *KIF11* (Extended Data Fig. 11e). As in the previous cases, the CoREs did not include the gene sequence but were either upstream or downstream of it. In both cases, CoREs changed from A to B subcompartments in tumours, and this repositioning was concomitant with increased H3K27me3 levels (Fig. 5j and Extended Data Fig. 11e) and loss of chromatin interactions with *XRCC5* and *KIF11* promoters (Fig. 5k and Extended Data Fig. 11f).

We noted that subcompartment repositioning events involving *CTNNB1* and *XRCC5* could be traced back to more moderate but concordant subcompartment changes already occurring in WGD cells (Fig. 5h,j). Notably, subcompartment changes detectable in WGD cells were concordant for 78–82% of the CoREs (termed consistent CoREs), frequently following a monotonic trajectory towards a more active or inactive compartment (Fig. 5l). These results were confirmed using an independent approach to select CoREs (Extended Data Fig. 11g,h). Overall, LCS initiates subcompartment changes that can result in CoREs, which leads to the deregulation of oncogenes and tumour suppressors independently of genetic alterations.

Tracing subcompartment changes in CP-A *TP53* $^{-/-}$ cells

To confirm our results in an independent model and experiments, we followed chromatin evolution in a subset of CP-A *TP53*$^{-/-}$ cells that spontaneously acquired high ploidy (Extended Data Fig. 12a), which suggested that they underwent WGD, and in CP-A *TP53*$^{-/-}$ clones in which WGD was induced (Fig. 6a). In the extremely small high ploidy cell population, we detected new translocations and compartment repositioning events (Extended Data Fig. 12b–e). However, in this model, cells probably underwent WGD at different time points and it was not possible to determine the timing of these events. Conversely, CP-A *TP53*$^{-/-}$ cells in which WGD was synchronously induced exhibited only minor compartment changes (Extended Data Fig. 12d,f) and gradual aneuploidization at 6-weeks and 20-weeks post-WGD (Supplementary Fig. 4). As these cells did not engraft in immunocompromised animals, we used the soft-agar assay to determine malignant transformation by assessing colony formation. Both clones 3 and 19 were able to form colonies post-WGD, and the colony size increased over time (Fig. 6b). By contrast, no colonies (clone 19) or only a limited number of small colonies (clone 3) were detectable in cells that did not undergo WGD (Fig. 6b).

Fig. 6: Copy number alterations and subcompartment repositioning in CP-A *TP53*$^{-/-}$ soft-agar colonies.



a, Timeline of in vitro and soft-agar colony-formation assay for CP-A *TP53*$^{-/-}$ post-WGD cells. **b**, Images of crystal violet colony staining and representative images of individual colonies for CP-A *TP53*$^{-/-}$ clone 3 and clone 19 ($n = 3$ independent wells). **c**, Volcano plots of segmented genomic regions, between each colony (col1 and col2) and control cells. *P* values calculated using DiffComp. **d**, Compartment rank in control (left), WGD (centre), and colonies (right) for each activating and inactivating CoRE. Lines connect compartment ranks belonging to the same CoRE. **e**, Overlap between CoREs identified in colonies derived from the same CP-A *TP53*$^{-/-}$ clone when considering all or only consistent CoREs. Grey bars represent

CoREs specific to one of the two colonies, whereas red bars denote common CoREs. **f**, Schematic representation of loss of chromatin segregation and subcompartment repositioning induced by WGD.

Source data

Next we performed Hi-C on 4 large colonies (2 from clone 3, 6 weeks post-WGD, and 2 from clone 19, 20 weeks post-WGD) and inferred CNVs and chromatin conformation changes. All colonies exhibited CNVs (Extended Data Fig. 12g) and CoREs (Fig. 6c), which were more similar among colonies derived from the same clone. Notably, 70–90% of the CoREs could be traced back to moderate but consistent compartment changes occurring in CP-A *TP53*$^{-/-}$ WGD cells (consistent CoREs) (Fig. 6d and Extended Data Fig. 12h). The overlap between CoREs found in two colonies derived from the same clone was higher when only consistent CoREs were considered (Fig. 6e). This result indicates that these were early events that emerged and were shared by most of the cells before they were transferred in soft agar. In summary, our results show that WGD induces both CIN and LCS that lead to the emergence of chromosomal alterations and subcompartment repositioning, which ultimately favour the selection of oncogenic epigenetic and transcriptional changes (Fig. 6f).

Discussion

Here we showed that WGD predisposes to the acquisition of a malignant phenotype, not only because of the emergence of CIN but also because of the reduced segregation of chromatin structural elements such as TADs and compartments. Increased contacts between usually well-segregated subcompartments culminate in subcompartment repositioning and epigenetic changes that support the activation of oncogenic transcriptional programmes.

However, to fully characterize the dynamic acquisition and selection of tumorigenic alterations, high-throughput and longitudinal single-cell molecular profiles are required. For example, it is tempting to speculate that increased contact frequency between chromosomes 1 and 16 in WGD cells (Fig. 3c) favoured the emergence of the translocation later observed in

tumours (Fig. 4g). More generally, it will be interesting to explore whether, similar to chromosomal alterations, heterogeneous chromatin 3D organizations exist at early time points after WGD and lead to the selection of tumour-promoting chromosome interactions and compartment changes. To test these hypotheses, highly multiplexed scHi-C experiments are required, possibly paired with barcoding technologies[49,50] and computational approaches to infer and trace chromatin structural elements across multiple time points. Expanding the scope of scHi-C data and analyses will be important to understand the contribution of chromatin 3D heterogeneity in malignant transformation. Notably, evidence from previous studies[7] and our work indicates that the oncogenic transformation of tetraploid cells is linked to a protein shortage , and activation of the tetraploid checkpoint is essential in non-cancerous tetraploid cells to restore protein levels. However, these findings were the results of targeted experiments focused on specific proteins. Future studies should investigate protein changes after WGD induction in an unbiased manner to determine whether additional phenotypes in WGD cells can be explained by insufficient protein synthesis.

Overall, our study demonstrated that in parallel to CIN, WGD induces LCS, which primes genomic regions for compartment changes that are selected and/or stabilized in tumour cells and are accompanied by epigenetic and transcriptional changes. These results provide a new lens to investigate the role of WGD and chromatin evolution in oncogenesis and tumour progression.

Methods

Cell culture

hTERT-RPE-1 WT and hTERT RPE-1 *TP53*$^{-/-}$ (46, XX)[27] cells were a gift from J. Korbel. The cells were grown in DMEM/F-12, GlutaMAX (10565018) supplemented with 10% FBS (Thermo Fisher Scientific, 10270106) and 1% antibiotic–antimycotic (Thermo Fisher Scientific, 15240062). CP-A (KR-42421) (47, XY) cells were purchased from the American Type Culture Collection (CRL-4027). CP-A *TP53*$^{-/-}$ cells were generated in this study using a CRISPR–Cas9 approach. The cells were

grown in MCDB-153 medium (Sigma-Aldrich, M7403) supplemented with 20 mg l^{-1} adenine (Sigma-Aldrich, A2786), 400 µg l^{-1} hydrocortisone (Sigma-Aldrich, H0135), 50 mg l^{-1} bovine pituitary extract (Thermo Fisher Scientific, 13028014), 1× insulin-transferrin-sodium selenite media supplement (Sigma-Aldrich, I1884), 8.4 µg l^{-1} cholera toxin (Sigma-Aldrich, C8052), 4 mM glutamine (Sigma-Aldrich, G7513), 5% FBS and 1% antibiotic–antimycotic. K562 (67, XX) cells were purchased from DSMZ (ACC 10) and cultured in RPMI medium (Thermo Fisher Scientific, 11875093) supplemented with 10% FBS and 1% penicillin–streptomycin (Thermo Fisher Scientific, 15140122). All cell lines were grown in a sterile, humidified incubator at 37 °C with 5% CO_2 and passaged every 3–5 days, depending on the cell line, to maintain appropriate cell densities.

Mice

All animals used in the study were NOD SCID gamma (NSG) female mice maintained at the EPFL animal facilities. Mice were kept in a 12 h-light 12 h-dark cycle, at 18–23 °C with 40–60% humidity, as recommended and in accordance with the regulations of the Animal Welfare Act (SR 455) and Animal Welfare Ordinance (SR 455.1). Mice were subcutaneously injected with 5 million cells in a 2:1 ratio of cell mixture to Matrigel basement membrane matrix (Corning, 354234), and tumour growth was monitored. Animal experiments were performed in accordance with the Swiss Federal Veterinary Office guidelines and as authorized by the Cantonal Veterinary Office (animal licence VD2932.1). Animals were sacrificed if the tumour volume was ≥1 cm^3.

Tissue dissociation

Subcutaneous tumours from mice were dissociated using a human tumour dissociation kit (Miltenyi Biotec, 130-095-929) with an enzyme cocktail and a gentleMACS dissociator with heaters (Miltenyi Biotec). The cell suspension was then strained through a 40 µm cell strainer (Corning, 352340). Samples were treated with 1× Red Blood Cell Lysis solution (Miltenyi Biotec, 130-094-183) for 10 min at 4 °C, and then spun down at 300*g* for 5 min and resuspended in 0.5% BSA in PBS. Last, mouse cells

were removed from the sample using a Mouse Cell Depletion kit (Miltenyi Biotec, 130-104-694) following the manufacturer's protocol.

CRISPR cloning

The sgRNA sequences targeting *TP53* (ref. [27]) were cloned into the pSpCas9(BB)-2A-GFP vector (PX458), which was a gift from F. Zhang (Addgene plasmid 48138; http://n2t.net/addgene:48138; RRID:Addgene_48138). In brief, 10 μM final concentration of each forward and reverse oligonucleotide were annealed in 1× T4 ligation buffer (New England Biolabs, B0202S) at 37 °C for 30 min, heated up at 95 °C for 5 min and ramped down by 0.1 °C s^{-1} to room temperature. In parallel, 10 μg of PX458 vector was digested with 10 U of BsmBI (New England Biolabs, R0580L) in 1× NEBuffer 3.1 at 55 °C for 1 h. Digested plasmid was run on a 1% agarose gel, extracted and purified using NucleoSpin Gel and PCR Clean-up (Macherey-Nagel, 740609) following the manufacturer's instructions. Annealed CRISPRs and digested plasmid were ligated using 5 U T4 DNA ligase (Thermo Fisher Scientific, EL0011) in 1× T4 DNA ligase buffer for 10 min at room temperature. The plasmid was then added to DH5α chemically competent bacteria and kept for 30 min on ice, followed by heat shock at 42 °C for 45 s. The bacteria were cooled down on ice and recovered in SOC medium for 1 h at 37 °C. Transformed bacteria were grown on ampicillin-containing growth medium at 37 °C overnight. Bacterial colonies were picked and expanded in LB broth supplemented with ampicillin for 12 h at 37 °C. Plasmid DNA was extracted using a Plasmid Plus Midi kit (Qiagen, 12945) according to the manufacturer's protocol. The plasmids were verified by Sanger sequencing (Microsynth) with hU6 primers.

CP-A *TP53* $^{-/-}$ cell line generation

CP-A WT cells were grown to 60–70% confluency in 10 cm plates. Next, 5 μg of PX458 plasmid containing *TP53*-targeting sgRNAs were diluted into 200 μl Opti-MEM I reduced serum medium (Thermo Fisher Scientific, 31985062). Then 15 μl FuGENE HD transfection reagent (for a 3:1 transfection reagent:DNA ratio) (Promega, E2312) was added to the DNA and incubated at room temperature for 15 min. The mixture was then added to a plate drop-by-drop and mixed by shaking. Cells were incubated for 48 h

at 37 °C in a humidified incubator. Transfected cells, expressing GFP, were single-cell sorted on a BD FACSAria Fusion instrument (BD Biosciences). Clones were allowed to expand, then individually tested by immunoblotting for TP53 protein levels following 24 h of treatment with 3 μM doxorubicin (Cayman Chemical, 15007).

WGD induction

Cells were seeded to 60–70% density. For mitotic slippage induction, 0.1 μg ml^{-1} nocodazole (Sigma-Aldrich, M1404) was added to the growth medium and CP-A and K562 cells were incubated for 72 and 48 h, respectively. For cells with an elongated G1 phase after tetraploidization, WGD in CP-A *TP53*$^{-/-}$ cells was induced with 0.1 μg ml^{-1} nocodazole for 72 h and treated with 0.5 μM of the CDK4/6i palbociclib (Sigma-Aldrich, PZ0383) for the last 16 h of the WGD induction protocol. For cytokinesis failure inductions, RPE cells were incubated for 24 h with 0.1 μg ml^{-1} nocodazole-containing medium. Following nocodazole treatment, the cells were exposed for an additional 24 h to 4 μM dihydrocytochalasin B (Cayman Chemical, 20845). The treatment was removed and cells were allowed to recover for 48 h to allow transition from a binucleated to a mononucleated state. Alternatively, WGD was induced through cytokinesis failure in RPE *TP53*$^{-/-}$ cells by incubation for 24 h with 9 μM of the CDK1 inhibitor RO-3306 (Sigma-Aldrich, SML0569) for G2 synchronization. The compound was washed off and cells were then treated with 4 μM dihydrocytochalasin B for 24 h. Cells were allowed to recover for 48 h, and tetraploid cells were sorted on the basis of cell cycle staining with 1 μg ml^{-1} Hoechst 33342 (Thermo Fisher Scientific, H1399).

Isolation of spontaneous high-ploidy cells

CP-A *TP53*$^{-/-}$ cells were stained with 1 μg ml^{-1} Hoechst 33342 for cell cycle profiling. Dividing cells with high ploidy (high Hoechst 33342 signal, >4N peak) were bulk sorted. Cells were allowed to recover overnight and then fixed for downstream analyses.

CIN induction

CIN was induced in RPE *TP53*$^{-/-}$ cells using a modified protocol described previously[51]. In brief, the cells were synchronized at the G1/S border with 5 mM thymidine (Sigma-Aldrich, T9250) for 24 h. Six hours after thymidine block release, the cells were treated with 500 nM of the MPS1 inhibitor reversine[52] (Sigma-Aldrich, R3904) for 12 h. Before processing for downstream analyses, cells were allowed to recover for 6 h.

Cell cycle staining

Cells were collected and washed with PBS (Thermo Fisher Scientific, 10010023). Permeabilization was performed in 0.01% Triton X-100 (AppliChem, A1388) in PBS for 30–60 min at 4 °C. Following PBS washes, the cells were fixed and stained with FxCycle PI/RNase staining solution (Thermo Fisher Scientific, F10797) overnight at 4 °C in the absence of light. Propidium iodide intensity for cell cycle detection was measured using Guava easyCyte (Luminex) and Galios (Beckman Coulter) cytometers and analysed using FlowJo (v.10.8) (BD).

Karyotyping

Cells were treated with 20 ng ml^{-1} KaryoMAX colcemid solution (Thermo Fisher Scientific, 15212012) for 2 h at 37 °C in a humidified incubator. Cells were collected in 0.8% sodium citrate solution (Sigma-Aldrich, S4641) and maintained at 37 °C for 30 min. The cell suspension was fixed with 3:1 methanol:acetic acid (Chemie Brunschwig, M/4000/17; FSHA/0406/PB08) added drop-by-drop, washed twice in the fixative solution and incubated overnight at −20 °C. Cells were dropped onto a glass slide (Thermo Fisher Scientific, J1800AMNZ). Slides were incubated for 2 min in a humidified chamber at 65 °C and air-dried at room temperature for 30 min. Slides were mounted and DAPI-stained concomitantly with ProLong Diamond antifade mountant with DAPI (Thermo Fisher Scientific,. P36962) according to the manufacturer’s instructions. Metaphases were imaged at ×100 resolution on a Zeiss Axioplan upright microscope. Images were analysed using Fiji (v.2.9.0)[53].

Immunoblotting

For non-histone proteins, cells were incubated in RIPA buffer consisting of 50 mM Tris-HCl pH 8.0, 1 mM EDTA, 1% Triton X-100, 0.5% sodium deoxycholate (Sigma-Aldrich, D6750), 0.1% SDS and 150 mM NaCl, for 30 min on ice for protein extraction. For histone extraction, cells were initially incubated with PBS lysis buffer consisting of 1% Triton X-100, 1 mM DTT (AppliChem, A2948), 1× protease inhibitor cocktail for 15 min at 4 °C and spun down at 12,000*g*. The resulting pellet was incubated overnight with 0.2 N hydrochloric acid (AppliChem, A5634). Lysates were then centrifuged at 12,000*g* for 10 min at 4 °C. Supernatant containing the protein fraction was isolated and mixed with 6× Laemmli sample buffer (12% SDS w/v, 60 mM Tris-HCl, pH 6.8, 50% glycerol (Fisher Scientific, G/0650), 600 mM DTT and 0.06% bromophenol blue (Sigma-Aldrich, B5525)) at 96 °C for 5 min. Mid-molecular weight proteins and histones were then separated on 12% or 15% SDS–PAGE gels, respectively, whereas high-molecular weight proteins were separated on a 7.5% Mini-PROTEAN TGX precast protein gel (Bio-Rad, 4561023). All gels were transferred onto 0.2 μm nitrocellulose membranes (Bio-Rad, 1704270) using a Trans-Blot Turbo transfer system (Bio-Rad, 1704150) according to the manufacturer's specifications. The membranes were blocked in a solution containing 5% milk (AppliChem, A0830) and 0.1% Tween-20 (Fisher Bioreagents, 10113103) in PBS for 30 min at room temperature. Blots were incubated in the same milk solution at either 4 °C overnight with primary antibodies against TP53 (Santa Cruz Biotechnology, sc-126; 1:500), β-actin (Cell Signaling Technology, 4967; 1:5,000), CTCF (Active Motif, 61311; 1:1,000), RAD21 (Abcam, ab992; 1:5,000) and α-actinin (Cell Signaling Technology, 6487; 1:1,000), or for 1 h at room temperature with primary antibodies against trimethyl-histone H3 (Lys9) (Cell Signaling Technology, 13969; 1:1,000), acetyl-histone H3 (Lys27) (Cell Signaling Technology, 8173; 1:1,000), trimethyl-histone H3 (Lys27) (Cell Signaling Technology, 9733; 1:1,000), and histone H3 (Cell Signaling Technology, 4499; 1:5,000). The membranes were incubated with fluorescent labelled goat anti-mouse (LI-COR Biosciences, 926-68070; 1:10,000) or goat anti-rabbit (LI-COR Biosciences, 926-32211; 1:10,000) for 2 h at room temperature and imaged using an Odyssey CLx imaging system (LI-COR Biosciences). Alternatively, the membranes were incubated with HRP-conjugated goat anti-mouse antibody (Merck, AP308P; 1:5,000) or goat anti-rabbit antibody (Merck, AP307P) for 1 h at room temperature. Blots were incubated with Amersham ECL western blotting detection

reagent (GE Healthcare, RPN2232) according to the manufacturer's instructions, and captured using a Fusion FX6 Edge imaging system (Witec). Images were analysed using Fiji (v.2.9.0)[53].

Immunofluorescence

Cells were cultured on coverslips coated with poly-d-lysine (Sigma-Aldrich, P7280) and incubated in standard conditions. Cells on coverslips were fixed with ice-cold methanol at 4 °C for 30 min. Cells were washed with PBS multiple times and incubated with 5% BSA (Sigma-Aldrich, A7906) at room temperature for 30 min. Next coverslips were incubated with primary antibodies at the indicated concentrations against pericentrin (0.1 μg ml^{-1}; Abcam, ab4448) and α-tubulin (0.5 μg ml^{-1}; Sigma-Aldrich, T6074) diluted in 1% BSA for 1 h at room temperature in a humidified chamber. Coverslips were washed with PBS and incubated with the fluorescent secondary antibodies anti-mouse IgG-Alexa Fluor 594 (2 μg ml^{-1}; Thermo Fisher Scientific, A-11005) and anti-rabbit IgG-Alexa Fluor 488 (2 μg ml^{-1}; Thermo Fisher Scientific, A-11034) diluted in 1% BSA for 1 h at room temperature in the dark. Coverslips were washed in PBS followed by mounting and counterstaining with DAPI with ProLong Diamond antifade mountant. Cell images were captured at ×63 resolution on a Zeiss Axioplan upright microscope. Images were analysed using Fiji (v.2.9.0)[53].

Soft-agar assay

For each condition, 100,000 cells resuspended in complete MDCB-153 medium were mixed in a 1:1 ratio with 0.7% sterile noble agar (Thermo Fisher Scientific, J10907). Cells were plated on Costar ultralow attachment plates (Corning, 3473) on top of a mixture of 1:1 MCDB-153 medium and 1.4% sterile noble agar. The mixture was allowed to solidify in a humidified atmosphere at 37 °C overnight, then fresh complete MCDB-153 medium was added on top of the layers of agar. Samples were incubated for up to 10 weeks in normal conditions, and the medium was replaced twice a week. Individual colonies were picked from the agar layer and cultured for downstream analysis. Last, the plates were stained with a solution of 0.5% crystal violet (Thermo Fisher Scientific, 405830250) in 20% ethanol for 30 min, washed with PBS and imaged.

Hi-C library preparation and analysis

Hi-C library preparation

Bulk Hi-C library preparation was performed as previously described[11,22], with minor modifications. Around 1–2 million cells were collected and fixed with 2% formaldehyde (Thermo Fisher Scientific, 11483217). The reaction was quenched with 200 mM glycine (VWR, 101194M) final concentration, cells were washed with PBS (Thermo Fisher Scientific, 10010023) and lysed in a solution containing 10 mM Tris-HCl pH 8.0 (Thermo Fisher Scientific, 15568025), 10 mM NaCl (Sigma-Aldrich, S6546), 0.2% IGEPAL CA-630 (Sigma-Aldrich, I8896) and 1× proteinase inhibitor cocktail (Roche, 11697498001) at 4 °C for 30 min. Resulting nuclei were resuspended in 1× NEB3.1 buffer (New England Biolabs, B7203S). The suspension of nuclei was incubated with 0.11% SDS (Carl Roth, CN30) final concentration for 10 min at 65 °C. The reaction was quenched with 1% Triton X-100 (AppliChem, A1388), and nuclei were digested with 100 U MboI restriction enzyme (New England Biolabs, R0147) at 37 °C overnight. The restriction enzyme was inactivated according to the manufacturer's specifications, and digested nuclei were washed and resuspended in 1× NEB3.1. Digested ends were then marked with biotin through incubation in 0.03 mM biotin-14-dATP (Thermo Fisher Scientific, 19524016), 0.03 mM dCTP, 0.03 mM dGTP, 0.03 mM dTTP (Promega, U1420) and 50 U Klenow DNA polymerase I (New England Biolabs, M0210M) for 4 h at room temperature. Resulting blunt-ends were proximally ligated with 50 U T4 DNA ligase (Thermo Fisher Scientific, EL0011), 1× T4 DNA ligase buffer (Thermo Fisher Scientific, B69), 5% PEG, 1% Triton X-100 and 0.1 mg ml^{-1} BSA (New England Biolabs, B9000S) for 4 h at room temperature. Crosslink reversal was performed on the proximity-ligated chromatin through incubation with 300 mM NaCl and 1% SDS overnight at 68 °C. The sample was then treated with 50 μg ml^{-1} RNase A (Thermo Fisher Scientific, EN0531) for 30 min at 37 °C, followed by 400 μg ml^{-1} proteinase K (Promega, V3021) at 65 °C for 1 h. DNA was purified by precipitation with 1.6 volumes of pure ethanol and 0.1 volumes of sodium acetate, pH 5.2 (Thermo Fisher Scientific, R1181) at −80 °C. DNA was eluted and then fragmented by sonication at 80 V peak incidence power, 10% duty factor, 200 cycles per burst for 60–80 s with an E220 focused-ultrasonicator

(Covaris). Sheared DNA was size-selected for library preparation using AMPure XP beads (Beckman Coulter, A63881). Next, biotin-marked fragments were isolated using Dynabeads MyOne Streptavidin C1 (Thermo Fisher Scientific, 65001), and all subsequent steps were performed on the bead-bound DNA fraction. Hi-C library preparation continued with an end polishing reaction, which involved incubation of DNA with 1× T4 ligase buffer (New England Biolabs, B0202S), 2.5 mM each dNTP, 50 U T4 polynucleotide kinase (New England Biolabs, M0201), 12 U T4 DNA polymerase (New England Biolabs, M0203) and 5 U Klenow DNA polymerase I, at room temperature for 30 min. PolyA tail was added by incubating the DNA sample in 1× NEBuffer 2 (New England Biolabs, B7002S) with 0.5 mM dATP and 25 U Klenow fragment (3′→5′ exonuclease) (New England Biolabs, M0212) at 37 °C for 30 min. DNA fragment ends were then ligated to Illumina TruSeq unique dual indexes (Integrated DNA Technologies) in 1× T4 ligation buffer with 5% PEG and 15 U T4 DNA ligase for 2 h at room temperature or overnight at 16 °C. Last, libraries were PCR amplified using Illumina forward (AATGATACGGCGACCACCGAGATCTACAC) and reverse (CAAGCAGAAGACGGCATACGAGAT) primers and KAPA HiFi HotStart ReadyMix (Roche, KK2602) for 6–10 cycles. Resulting fragments were size-selected using AMPure XP beads. Libraries were sequenced in a PE150 configuration on HiSeq X, NovaSeq 6000 or HiSeq 2500 systems (Illumina).

Generation of Hi-C contact maps

For each library replicate, reads were mapped to the human hg19 reference genome using bwa mem (v.0.7.17)[54] and processed using the Juicer pipeline (v.1.6)[55]. For each sample, Hi-C maps were generated at the following resolutions: 10 kb, 20 kb, 25 kb, 50 kb, 100 kb, 250 kb, 500 kb, 1 Mb and 10 Mb. Once the concordance between replicates of the same biological condition was assessed, Hi-C maps of the same condition were merged using the mega.sh script provided in the Juicer pipeline. All Hi-C maps were normalized using the Knight–Ruiz method (KR)[56] implemented in the Juicer pipeline.

Definition of Hi-C compartments

Compartments were called using the Calder pipeline[29] on KR-normalized Hi-C maps at 50 kb resolution. Calder returns a segmentation of the genome in compartments where each segment is assigned both a compartment rank (a real number between 0 and 1) and a compartment label (B.2.2, B.2.1, …, A.1.2, A.1.1), which is a discretization of the rank in eight different categories. Compartment ranks and labels correlate with the chromatin state of the DNA region, with values close to 0 being more B-like compartments and values close to 1 being more A-like compartments.

Assessment of similarity between Hi-C contact maps

Pairwise comparisons between intrachromosomal contact maps were based on the following metrics: a correlation measure between the contacts, stratified by the distance between the interacting loci; the conservation of compartment domains and their boundaries; the correlation at the level of boundary insulation; and the correlation at the level of Calder compartment rank.

Replicates of the same biological condition (control versus control, WGD versus WGD) and samples of different conditions (control versus WGD, control vs 20-weeks post-WGD tumours) were compared, as well as samples of a different cell line (control versus GM12878 from ref. [12]). Inter-replicate comparisons and intercell line comparisons gave a reference baseline of random fluctuations and extensive chromatin changes, respectively, for each score.

Correlation of contacts (stratum-adjusted correlation coefficient)

The stratum-adjusted correlation coefficient[57] was used as implemented in the HiCRep.py package[58]. The maximum genomic distance to test was set to 10 Mb, and Hi-C maps were binned at 100 kb and smoothed with a window of $H = 3$ bins. For each comparison, a stratum-adjusted correlation coefficient value was computed for each chromosome.

Conservation of compartment domains

Compartment domains were called at 50 kb resolution using the Calder pipeline on the KR-normalized Hi-C matrices.

Given two compartment domain sets identified on the same chromosome in two samples, the measure of concordance[59] was calculated, which was previously defined to compare two clustering assignments. The measure of concordance is a real number bounded between 0 and 1, with 1 representing identical chromosome segmentation and 0 maximum discordance.

Conservation of insulating boundaries

Hi-C insulation was computed as previously described[30]. Insulation scores for each chromosome were calculated using the FANC library[60], specifically, the InsulationScores.from_hic function on the KR-normalized intrachromosomal Hi-C matrices at 50 kb resolution using a sliding window of 1 Mb. Sliding windows with more than 20% of missing values were not considered. Scores were normalized by the geometric mean chromosome-wise and finally $\log_2$-scaled. The final score is therefore centred at 0, with local minima representing putatively TAD boundaries.

Comparisons between samples were performed by computing the Spearman correlation coefficient of the insulation scores for each chromosome.

Hi-C compartment similarity

The compartment segmentation given by Calder was split in bins of 50 kb, assigning to each bin the compartment rank of the segment it belongs to. The similarity between two samples was then computed separately for each chromosome as the Spearman correlation of the two binned rank vectors.

Hi-C interchromosomal similarity

The interchromosomal interactions for each pair of Hi-C maps were compared by considering separately the interactions between each pair of different chromosomes in the two samples. The Spearman correlation coefficient of the raw interaction counts was computed between the two

samples for each chromosome pair. For each Hi-C comparison, therefore, a correlation value for each pair of chromosomes was obtained.

Analysis of Hi-C interchromosomal interactions

To determine interaction biases between pairs of chromosomes, Hi-C interactions were aggregated between each pair of chromosomes, obtaining a 23 × 23 interaction matrix I. The matrix was balanced using iterative correction[61] to remove interaction biases due to the length of the chromosomes (such that the marginal sum of each chromosome is 1). This resulted in a normalized matrix I_{ICE}. This normalization is similar to the one presented in ref. [11], with the advantage of ensuring constant marginals. When compared, both normalizations produced comparable results.

Chromosomes were then divided into two clusters on the basis of their interaction profile in I_{ICE}: chromosomes from 1 to 14 and X were categorized as long, whereas chromosomes from 15 to 22 were categorized as short.

To compare control and WGD interchromosomal interaction matrices, their ratio $R = \log_2[I_{\text{ICE}}(\text{WGD})/I_{\text{ICE}}(\text{Control})]$ was computed. Chromosome interactions were then split into three categories on the basis of the chromosome cluster of their ends: long–long, long–short and short–short. Chromosome interaction categories were compared by computing a Mann–Whitney test P value between R values of each pair of categories.

Hi-C interchromosomal map balancing at 10 Mb resolution

To visualize interchromosomal Hi-C maps at 10 Mb resolution, Iterative Correction using the Cooler package[62] was performed. Counts were normalized such that each bin had total number of interchromosomal interactions equal to 1.

Analysis of Hi-C intercompartmental interactions

The genomic segmentation in eight classes given by Calder (B.2.2 to A.1.1) was considered. Each 50 kb genomic bin was then associated to the compartment level it belongs. For each chromosome, its intrachromosomal contacts were extracted at 1 Mb resolution and then normalized by genomic distance[12] using the FANC package[60]. These interactions were then upscaled to 50 kb resolution by assigning to each 50 × 50 kb pixel the value of the 1 × 1-Mb superpixel it belonged to. This procedure was performed to smooth the normalized interaction values and to ensure enough coverage for each genomic distance. For each 50 kb genomic bin b, the sum of the normalized interactions between that bin and the bins belonging to the eight compartment level classes was computed separately, thus obtaining one value s_b(comp) for each compartment level comp. These values were then divided by the total sum of interactions of that bin T_b. To consider the bias induced by the amount of chromosome covered by each compartment level, these values were further divided by the percentage of bins belonging to each compartment level $B_{\rm comp}$, thus obtaining $z_b({\rm comp}) = s_b({\rm comp})/T_b B_{\rm comp}$. The obtained value was finally divided by their sum to obtain for each bin $f_b({\rm comp}) = z_b({\rm comp})/\Sigma_c z_b(c)$, which is a number between 0 and 1 for each compartment level representing the level of segregation of each compartment level for that bin. For each bin b, it was defined as $C{\rm Score}_b$ the segregation level of the compartment level to which the bin belongs to. This definition is an adaptation of the compartment score computed in ref. [63], but applied at the bin level.

Given two conditions, for example, WGD and control, the difference for each pair of subcompartments $\rm comp_1$, $\rm comp_2$ was computed as follows:

$$\begin{array}{l}{\sigma }_{{{\rm{comp}}}_{1}}({{\rm{comp}}}_{2})=\frac{{\rm{No.\; times}}\,{f}_{b}^{{\rm{WGD}}}({{\rm{comp}}}_{2}) < {f}_{b}^{{\rm{Control}}}({{\rm{comp}}}_{2})}{{\rm{No.\; times}}\,{f}_{b}^{{\rm{WGD}}}({{\rm{comp}}}_{2})\ge {f}_{b}^{{\rm{Control}}}({{\rm{comp}}}_{2})},\\ \,\,\,\,{\rm{for}}\,b\in {{\rm{comp}}}_{1}\end{array}$$

and

$$\sigma \left({{\rm{comp}}}_{1},{{\rm{comp}}}_{2}\right)=-{\log }_{2}\left(\frac{1}{2}\left({\sigma }_{{{\rm{comp}}}_{1}}\left({{\rm{comp}}}_{2}\right)+{\sigma }_{{{\rm{comp}}}_{2}}\left({{\rm{comp}}}_{1}\right)\right)\right).$$

${\sigma }_{{{\rm{comp}}}_{1}}\left({{\rm{comp}}}_{2}\right)$ represents the ratio between the number of bins in $comp_1$, which lose compartment segregation with $comp_2$, and the number of bins in $comp_1$, which gain compartment segregation with $comp_2$.

$\sigma \left({{\rm{comp}}}_{1},{{\rm{comp}}}_{2}\right)$ is simply the average of ${\sigma }_{{{\rm{comp}}}_{1}}\left({{\rm{comp}}}_{2}\right)$ and ${\sigma }_{{{\rm{comp}}}_{2}}\left({{\rm{comp}}}_{1}\right)$, which makes it a symmetric measurement of average segregation changes between compartment levels $comp_1$ and $comp_2$. The $-\log_2$ of this number was computed for representation purposes, with positive and negative $\log_2$ ratios indicating gain and loss of contacts, respectively, between the two compartment levels.

To specifically assess the extent of loss of segregation for a specific compartment level comp, a similar strategy was adopted:

$${\sigma }_{{\rm{comp}}}=\frac{{\rm{No.\; times}}\,C{{\rm{Score}}}_{b}^{{\rm{WGD}}} < C{{\rm{Score}}}_{b}^{{\rm{Control}}}}{{\rm{No.\; times}}\,C{{\rm{Score}}}_{b}^{{\rm{WGD}}}\ge C{{\rm{Score}}}_{b}^{{\rm{Control}}}},{\rm{for}}\,b\in {\rm{comp}}$$

was computed, which represents the ratio between the number of bins in comp losing segregation and the number of bins in comp gaining segregation. Values higher than 1 indicate loss of segregation, whereas values below 1 indicate gain of segregation.

Boundary insulation analysis

TAD boundaries in control and WGD samples were determined from insulation scores using the fanc.Boundaries.from_insulation_score function from the FANC package[60], looking at local minima of the score in the 400 kb region around the bin. Each boundary was assigned the insulation score corresponding to its position. Lower values of the score signify higher insulation capability of the boundary. The boundaries shared between the two samples (±50 kb) were then extracted. The top 300 insulating boundaries were selected as follows: control and WGD boundaries were separately ranked on the basis of their insulation scores. For each condition, the top 300 ranked boundaries were selected and their maximum insulation score (corresponding to the weaker boundary in the set) was determined, which was called $I_{{\rm{top}}300}^{{\rm{Control}}}$ and $I_{{\rm{top}}300}^{{\rm{WGD}}}$, respectively. An insulation threshold $I_{{\rm{top}}300}={\rm{\max}}\left(I_{{\rm{top}}300}^{{\rm{Control}}}, I_{{\rm{top}}300}^{{\rm{WGD}}}\right)$ was defined. Finally, shared boundaries between control and WGD having insulation scores smaller than $I_{{\rm{top}}300}$ were selected. It should be noted that this approach does not ensure that the final number of selected boundaries is exactly 300.

Independence of LCS measurement from Hi-C resolution and coverage per haploid copy

The aggregated map of RPE *TP53*$^{-/-}$ WGD cells (218 million reads) was compared with one of the control replicates maps (108 million reads). Conversely, one replicate of the control (108 million reads) was compared with the aggregated map of the same control (221 million reads).

Detection of regions of significant CoREs

To determine significant CoREs, we developed an algorithm to identify contiguous genomic regions with consistently different compartment ranks computed using Calder. We refer to this method as DiffComp. A segmentation algorithm was designed as follows. Given two Hi-C experiments *X* and *Y*, the genomic segmentations of both in compartment domains was determined using Calder on the 50 kb resolution KR-normalized Hi-C matrices. Both segmentations were then binned in 50 kb

bins, assigning to each bin its relative compartment rank. Thus, for each chromosome, compartmentalization in the two samples is represented as two numerical vectors $\mathbf{C}_X$, $\mathbf{C}_Y$.

The pairwise rank difference for each genomic bin were computed as $\Delta\mathbf{R}_{XY} = \mathbf{C}_X - \mathbf{C}_Y$. This vector represents the differential rank between the two experiments, with positive values indicating a shift towards active compartments and negative values indicating a shift towards inactive compartments.

The genome was segmented based on $\Delta\mathbf{R}_{XY}$ using a recursive strategy. Given σ^*, which represents the maximum allowed standard deviation in the signal that a segment can have before being split into subsegments, the procedure involves the following process.

Each chromosome is initially considered a single whole segment and then

1. (1)

   The standard deviation of the segment $\sigma(s)$ and its average value mean*(s)* were calculated.

2. (2)

   If $\sigma(s) < \sigma^*$, then the procedure stops and the segment is assigned mean*(s)* as value, which represents its subcompartment repositioning score.

3. (3)

   Otherwise, the segment is split into subsegments depending on whether they are above or below the mean*(s)* value.

4. (4)

   For each of the subsegments, the procedure is repeated from point (1).

The expected distribution of compartment changes can be computed using technical or biological replicates of the same experiment. An expected differential vector $\Delta\mathbf{R_E} = \mathbf{C}_{\mathrm{R1}} - \mathbf{C}_{\mathrm{R2}}$ was computed using two replicates of RPE *TP53*$^{-/-}$ control.

In this analysis, $\sigma^* = 0.1$ was fixed, which is 1.3-times the standard deviation of $\Delta\mathbf{R_E}$. For each detected compartment repositioning segment s, an empirical P value as P value$(s) = P(\max\{|\Delta\mathbf{R}_{XY}(s)|\} < |E|)$ was computed, where $E \in \Delta\mathbf{R_E}$. This P value depends both on the average value of the segment and on its length, for which longer segments have higher statistical power.

The output of this method is a list of CoRE regions together with their average compartment repositioning score (which can vary from −1 and 1) and their computed empirical P value.

For each comparison studied, CoRE regions having an absolute average value above 0.1, an empirical P value below 0.01 and a segment length above 300 kb were considered.

CoRE overlap in CP-A *TP53* $^{-/-}$ colonies

To assess the amount of overlap between two sets of CoREs C_1, C_2 belonging to different sample comparisons, the two sets were divided in activations and inactivations on the basis of the sign of the compartment repositioning score ($C_1^{\mathrm{A}}, C_1^{\mathrm{I}},\, C_2^{\mathrm{A}}, C_2^{\mathrm{I}}$). The CoREs of the same type coming from both sample comparisons ($C_{12}^{\mathrm{A}}, C_{12}^{\mathrm{I}}$) were merged together by stacking overlapping regions together (using the bedtools merge command from bedtools (v.2.30.0)[64,65]), finally creating a consensus set of CoREs concatenating the two sets ($C_{12} = [C_{12}^{\mathrm{A}}, C_{12}^{\mathrm{I}}\,]$). Each consensus CoRE was checked for overlapping with a CoRE of the same type in C_1 and C_2. Consistent CoREs were considered overlapping when two CoREs of the same type were overlapping and at least one of the two was a consistent CoRE.

Tracing compartment repositioning from WGD to tumour time points

For each of the tumours, the CoRE regions that passed the previously defined statistical filters were considered. For each of these regions s, the corresponding Calder segmentation in the WGD and control time points were extracted. The average compartment rank of the CoRE region in the two previous time points were then computed, which were defined as $r_{\mathrm{WGD}}(s)$ and $r_{\mathrm{Control}}(s)$, respectively. The compartment rank of the CoRE region in the tumour were defined as $r_{\mathrm{Tumour}}(s) = r_{\mathrm{Control}}(s) + \mathrm{mean}(s)$, where $\mathrm{mean}(s)$ is given by the CoRE detection algorithm. A parameter ε was then defined, which is the minimum absolute rank difference between WGD and control, namely $|r_{\mathrm{WGD}}(s) - r_{\mathrm{Control}}(s)|$, to classify the CoRE region as activating or inactivating in WGD with respect to control.

Given ε, CoRE regions in tumours can be discriminated on the basis of the type of change in tumours (activating or inactivating) as well as the type of change at WGD (unchanged, activating or inactivating). The number of CoRE regions belonging to each of the six combinations was counted.

A CoRE region was defined as ‘consistent’ when it belongs to the activation–activation or the inactivation–inactivation class. The percentage of consistent CoRE regions were counted with different choices of parameter ε. Observing the steepness of the curves in the three tumour samples, a shared parameter to $\varepsilon = 0.05$ was fixed.

Comparing different segmentation algorithms for CoRE detection

The segmentation strategy in DiffComp was compared to the circular binary segmentation (CBS) algorithm, which was previously developed for the segmentation of copy number changes[66]. CBS was applied to the Calder differential rank vector $\Delta\mathbf{R}_{XY} = \mathbf{C}_X - \boldsymbol{C}_Y$, which was computed as explained above. Segments detected using CBS were annotated with their compartment repositioning score and P values as described above. CoREs were then filtered on the basis of the repositioning score, P value and size as defined above. The CoREs detected using DiffComp and the ones detected with CBS were compared using as the benchmark the RPE *TP53*$^{-/-}$ 20 week

post-WGD tumour 1 versus RPE *TP53*$^{-/-}$ control comparison. We then compared the breakpoint positions between the two segmentation strategies, the corresponding sets of significant CoREs and the traceability of these events to subcompartment changes occurring in WGD cells.

Hi-C phasing in RPE *TP53* $^{-/-}$ 20-week post-WGD tumours

Integrated phasing was performed using Hi-C reads from the pooled RPE Hi-C replicates (control). Single-nucleotide variants (SNVs) were first identified from the Hi-C reads using Freebayes[67] (version v1.3.2-46-g2c1e395-dirty). SNVs were phased into two haplotypes, namely Hap1 and Hap2, using a previously described integrated phasing strategy[68]. In brief, population-based phasing was first conducted using SHAPEIT2 (ref. [69]; version v2.904.3.10.0-693.11.6.el7.x86_64) with hg19 1000 Genomes project phase 3 as a reference panel. Pseudo-reads generated from the population haplotype likelihood were then combined with the Hi-C reads as input to HapCUT2 (ref. [70]) for the second round of phasing. This approach returned several phasing blocks for each chromosome. Phasing information was retained only from the most variants phased block, which harbours the majority of input SNVs (>90%). Only Hi-C interactions for which anchors mapped strictly to one of the two haplotypes were retained for analysis, thus obtaining three sets of interaction types: Hap1–Hap1, Hap1–Hap2 and Hap2–Hap2.

Analysis of contacts between homologous chromosomes after WGD

After WGD, the rates both *in cis* and *in trans* contacts are expected to increase. In detail, putative *in cis* contacts should increase by a factor of 3 following WGD, and *in trans* contacts should increase by a factor of 4. Hence, it is expected that the ratio of *in trans* (T) versus *in cis* (C) contacts increases after WGD as described below:

$$\begin{array}{ccc}{r}_{{\rm{Control}}}=\frac{{\rm{T}}}{2{\rm{C}}}{\rm{;}} & {r}_{{\rm{WGD}}}=\frac{4{\rm{T}}}{6{\rm{C}}}{\rm{;}}

$$\frac{{r}_{{\rm{WGD}}}}{{r}_{{\rm{Control}}}}=\frac{4}{3}=1.33$$

To verify this prediction, *in cis* interactions were defined as all the Hi-C-phased interactions of the type Hap1–Hap1 and Hap2–Hap2, and *in trans* interactions all the Hi-C-phased interactions of type Hap1–Hap2. The following was computed:

$${r}_{{\rm{Control}}}=\frac{{\rm{No.}}\,{({\rm{Hap}}_{1}-{\rm{Hap}}_{2})}^{{\rm{Control}}}}{{\rm{No.}}\,{({\rm{Hap}}_{1}-{\rm{Hap}}_{1})}^{{\rm{Control}}}+{\rm{No.}}\,{({\rm{Hap}}_{2}-{\rm{Hap}}_{2})}^{{\rm{Control}}}}$$

$${r}_{{\rm{WGD}}}=\frac{{\rm{No.}}\,{({\rm{Hap}}_{1}-{\rm{Hap}}_{2})}^{{\rm{WGD}}}}{{\rm{No.}}\,{({\rm{Hap}}_{1}-{\rm{Hap}}_{1})}^{{\rm{WGD}}}+{\rm{No.}}\,{({\rm{Hap}}_{2}-{\rm{Hap}}_{2})}^{{\rm{WGD}}}}$$

and $\frac{{r}_{{\rm{WGD}}}}{{r}_{{\rm{Control}}}}$ separately for each chromosome, and for the genome-wide average ratio. Finally, $\frac{{r}_{{\rm{WGD}}}}{{r}_{{\rm{Control}}}}=1.25$ was obtained, close to the predicted value.

Calling copy number alterations from bulk and phased Hi-C reads

A strategy to impute broad CNVs from Hi-C data was designed as follows:

1. (1)

   For each bin b, its coverage n_b was computed.

2. (2)

   Bins overlapping genomic gaps and bins having ${n}_{b} < \bar{R}-\gamma M$ were excluded by the analysis, with $\bar{R}$ being the genome-wide coverage median, *M* being the median genome-wide absolute deviation of the coverage and $\gamma \in {\mathbb{N}}$ begin a defined parameter.

3. (3)

   n_b was normalized by the median chromosome coverage ($\bar{{R}_{{\rm{C}}}}$), obtaining $\widetilde{{n}_{b}}={n}_{b}/\bar{{R}_{{\rm{C}}}}$. This step enables to identify copy number changes at the subchromosomal level.

4. (4)

   The CBS algorithm[66] was run on $\widetilde{{n}_{b}}$. If a chromosome has no breakpoints, the entire chromosome is defined as a segment.

5. (5)

   For each segment *s*, the median value of its genome-wide normalized coverage, ${w}_{s}={\rm{median}}{({n}_{b}/\bar{R})}_{b\in s}$ was computed.

6. (6)

   CNVs were defined as follows: all segments or chromosomes having $\left|{w}_{s}-1\right|\ge t$, with *t* being a defined threshold representing the minimum absolute difference from the genome-wide median coverage a segment has to have to be defined as a CNV.

For bulk Hi-C data of the CP-A *TP53*$^{-/-}$ colonies, a bin size of 2 Mb was used, with $t = 0.4$ and $\gamma = 7$. For phased Hi-C data, $\gamma = 4$ was used.

Detecting significant interactions in RPE *TP53* $^{-/-}$ control cells and post-WGD tumours

HiC-DC[71] was used to compute the statistical significance of chromatin interactions at the bin level (20 kb resolution). The degree of freedom in the hurdle negative binomial regression model was set as 6. The sample size parameter was determined by trying 20 values in the (0.5,1) range with equal distance, then choosing the maximum value that did not result in optimization failure in R. Other parameters of HiC-DC were set as default.

RNA-seq protocol and analysis

RNA-seq library preparation

RNA was extracted from RPE *TP53*$^{-/-}$ control cells and WGD cells using a RNeasy Mini kit (Qiagen, 74104) following the manufacturer's protocol. Resulting RNA was processed for sequencing using a TruSeq Stranded mRNA kit (Illumina, 20020594) according to the supplier's recommendations. Libraries were then sequenced on an Illumina NovaSeq 6000 platform in a PE150 configuration.

RNA-seq data processing and analysis

RNA-seq fastq files were analysed using the nfcore/rnaseq pipeline (v.3.8; https://nf-co.re/rnaseq) using as the aligner star_rsem (ref. [72]), mapping the reads to the hg19 genome. Differentially expressed genes between WGD and control were determined using DESeq2 (ref. [73]). Genes having an absolute $\log_2$(FC) above 0.1 and a *P* value of <0.01 were considered significantly differentially expressed. Gene set enrichment analysis was performed using Enrichr[74].

Relationship between gene expression changes and LCS at WGD

Each gene was associated to the 50 kb bin containing its transcription start site. Each gene bin was then associated to the compartment rank computed

by Calder in RPE *TP53*$^{-/-}$ control and WGD and computed the difference (Δcompartment). To check the association with boundary insulation changes after WGD, the genes for which the transcription start site was in proximity of an insulation boundary in RPE *TP53*$^{-/-}$ control (±50 kb) having an insulation score below −0.1 were considered. The percentage of upregulated and downregulated genes in proximity of boundaries gaining and losing insulation and the fold changes against the percentages in the total set of genes were computed.

scHi-C protocol and analysis

scHi-C library preparation

Single-cell Hi-C was performed using a modified protocol described previously[75]. Fixation of the nuclei with formaldehyde, MboI digestion and biotin fill-in was performed in a pool of 1 million cells following the same procedure as described for bulk processing. Next in-nucleus proximity ligation was done with 50 U T4 DNA ligase, 1× T4 DNA ligase buffer, 5% PEG, 1% Triton X-100 and 0.1 mg ml^{-1} BSA at 16 °C overnight, with light mixing. Only pools of nuclei with at least 75% of the population showing an integral nuclear membrane were considered for further processing. Nuclei were strained multiple times through a 10 µm nylon net filter (Merck-Millipore, NY1009000). Sample preparation was done using DispenKit (SEED Biosciences), and single nuclei were dispensed in skirted Eppendorf twin.tec PCR plate 96-wells (Eppendorf, 0030128648) containing 50 µl of NEBuffer 3.1 (New England Biolabs, B7203S) using the single cell isolator DispenCell (SEED Biosciences) following the manufacturer's instructions. A nucleus passing through the tip of the DispenCell, which acts as a Coulter counter[76], leaves an electrical impedance change mark, which is proportional to the volume of the nucleus. For RPE *TP53*$^{-/-}$ cells, a minimum impedance change of 75 Ω was detected for the diploid nucleus and 200 Ω for the tetraploid nucleus. A lower impedance change was associated with debris or unsuccessful induction of genome doubling in the case of the WGD condition; thus, such nuclei were not considered for further processing. To avoid processing of nuclei aggregates, dispensed single nuclei associated with a threshold higher than 400 Ω and 800 Ω for

diploid and tetraploid nuclei, respectively, were also discarded. These ranges were set from quality metrics of a test scHi-C batch. Specifically, an unreasonable number of unique interactions of each fragment end (for example, >2 for diploid loci) would indicate a nuclei aggregate rather than a single nucleus. Following dispensing, the single nuclei were de-crosslinked by incubation at 65 °C overnight. Next each selected nucleus was mixed with 25 µl Dynabeads MyOne Streptavidin C1 and transferred to a 1.5 ml tube and incubated at room temperature for 1 h on a rotating wheel. The bead-bound fragments were digested with 10 U of AluI restriction enzyme (New England Biolabs, R0137) in 1× rCutSmart buffer (New England Biolabs, B6004) at 37 °C for 2 h. A-tailing reaction and adapter ligation were performed for each single cell as for the bulk Hi-C processing. Similarly, PCR amplification of the single cell libraries was performed in the same master mix as described above for bulk Hi-C, for 27–30 cycles. Libraries were cleaned using AMPure XP beads and then ran on a 2% agarose gel at 100 V for 50–60 min. Successful libraries presented a 300–700 bp smear, which was cut from the gel. DNA purification from the agarose was performed using NucleoSpin Gel and PCR clean-up (Macherey-Nagel, 740609) following the manufacturer's instructions. An additional AMPure XP bead size selection was generally necessary to remove any primer dimer contamination. Libraries were sequenced on a NextSeq550 platform (Illumina) in a PE75 configuration.

scHi-C contact map generation and quality filtering

For each cell, paired-end R1 and R2 fastq files were separately aligned to the hg19 reference genome using bwa mem (v.0.7.17). The scHiCExplorer pipeline[77] was used to generate Hi-C contact maps in Cooler format[62]. Quality control of scHi-C interactions was performed as previously reported[75,78]. Specifically, as it was previously reported that end-pairs covered by only one read are probably results of alignment or pairing errors of the sequencing machine[78], they were removed from the analysis. Cells for which the percentage of singleton interactions was above 75% were also removed from the analysis. Additionally, cells with fewer than 100,000 unfiltered interactions were removed. Finally, for the remaining end-pairs, which were supported by at least two duplicated reads, duplicates were removed. One cell was removed from the analysis because of the absence of

interchromosomal interactions involving chromosome 1, which indicated the occurrence of technical issues during the library preparation.

Analysis of single-cell Hi-C inter-chromosomal interactions

Similar to bulk Hi-C, for each cell *x* passing the previously defined filters, inter-chromosomal interactions between each pair of chromosomes were aggregated and the ICE-balanced interaction matrix $I_{\mathrm{ICE}}(x)$ was obtained. A loss of chromosomal segregation score was then defined for each cell $\mathrm{LCS}(x) = \mathrm{LS}(x)/(\mathrm{LL}(x) + \mathrm{SS}(x))$, where: $\mathrm{LS}(x)$ = the number of balanced interactions between long and short chromosomes; $\mathrm{LL}(x)$ = the number of balanced interactions between long and long chromosomes; and $\mathrm{SS}(x)$ = the number of balanced interactions between short and short chromosomes.

The higher the LCS score, the higher the loss of chromosomal segregation in the analysed cell. The scores were compared between the control and WGD RPE $TP53^{-/-}$ populations and a Wilcoxon two-tailed *P* value was computed.

Chromosome pairs having the highest interaction enrichment or depletion in the WGD population were identified by sorting chromosome pairs for each cell by their number of balanced interactions and then computing the average rank of each pair in the control and WGD populations.

Definition of scHi-C compartment segregation

A simplified strategy for compartment imputation was adopted. For each cell, the intrachromosomal contact matrices of each chromosome at 1 Mb resolution were extracted. Bins having zero marginal counts were removed from the analysis. The observed over expected matrix was then calculated as previously described[11], whereby the contact decay profile was computed in logarithmically increasing bins using the cooltools package[62]. The normalized matrix was centred around 0 by subtracting 1. Next, the Pearson correlation of each pair of bins was computed and, finally, the first two principal component analysis (PCA) components for each bin of the matrix were extracted. A and B compartments were assigned to each scHi-C bin associating the relative compartment in the bulk Hi-C by aggregating Calder segmentations of RPE $TP53^{-/-}$ control samples into A and B regions.

Segregation scores for single cells were then computed as the silhouette score between A and B clusters[79] for each chromosome, using the two previously determined PCA components as point coordinates in the two-dimensional cartesian plane.

Definition of single-cell compartment consistency across conditions

scHi-C bins were clustered on the basis of the two PCA components using the KMeans algorithm[80], with the number of clusters fixed to 2. Next the adjusted Rand index[81] between the *K*-means clusters and the bulk A and B clusters was computed. Only intrachromosomal maps with a score above 0 were considered for analysis. The two *K*-means clusters were renamed into A and B, such that the correlation with the bulk compartments was maximized. The consistency of compartment calls across cells of the same biological condition (control or WGD) was calculated for each 1 Mb bin as follows: if A and B are respectively the number of cells in which the bin was called as the A or B compartment, the consistency of the bin is max(A, B)/(A + B).

Pseudo-bulk scHi-C analysis

Interchromosomal pseudo-bulk Hi-C values were derived from individual cell interchromosomal counts by summing all interactions for each chromosome pair, thus obtaining a 23 × 23 chromosome interaction matrix *I*. Observed and expected (O/E) chromosome level interactions were derived as described in ref. [11]. In brief, the number of observed interactions between a pair of chromosomes I_{ij} was divided by the number of possible interactions between the two chromosomes E_{ij}. E_{ij} was empirically estimated as the product between the total number of interchromosomal interactions of the first chromosome (C_i) and the second chromosome (C_j), divided by double the total number of interchromosomal interactions (N):

$${E}_{ij}=\frac{{{\rm{C}}}_{i}\times {{\rm{C}}}_{j}}{2N}$$

A chromosome-level interaction enrichment matrix OE = *I/E* was obtained.

To remove noise, a correlation was computed for each chromosome pair (*i,j*) as the Spearman correlation of the vectors corresponding to the rows of the two chromosomes in the OE matrix (OE_i, OE_j). A 23 × 23 correlation matrix ρ was obtained such that ρ_{ij} = Spearman (OE_i, OE_j).

Differences in correlation between control and WGD in pseudo-bulk Hi-C were calculated as the $\log_2$ ratio of the correlations shifted by 1:

$${\sigma }_{i,j}={{\rm{\log }}}_{2}\left(\frac{{\rho }_{{ij}}^{{\rm{WGD}}}+1}{{\rho }_{{ij}}^{{\rm{Control}}}+1}\right)$$

Calling copy number alterations from scHi-C

The same procedure as defined for bulk and phased Hi-C data was performed, using as bin size of 5 Mb and a maximum absolute deviation threshold $t = 0.4$.

scRNA-seq and analysis

Sequencing

scRNA-seq was performed using a Chromium Next GEM Single Cell 3′ kit v.3.1 (10x Genomics) following the manufacturer's protocol. The number of cells targeted for each condition was 3,000. Resulting libraries were sequenced on a NovaSeq 6000 or HiSeq 4000 system (Illumina).

scRNA-seq read alignment and data processing

The sequencing reads for all the samples (RPE *TP53*$^{-/-}$ control, RPE *TP53*$^{-/-}$ 6-weeks post-WGD, RPE *TP53*$^{-/-}$ 20-weeks post-WGD, RPE *TP53*$^{-/-}$ 20-weeks post-WGD tumours T1–T3) were aligned using the human reference transcriptome (hg19, Ensembl-87 build) with the 10x Genomics Cell Ranger pipeline (v.3.1.1)[82] with default parameters. The Seurat R package (v.3.1.5)[83] was utilized for data processing. Raw unique molecular identifier (UMI) read count data for each sample were read as Seurat data objects by

keeping genes expressed in at least one cell. The following number of cells were acquired and retained after filtering for each sample: RPE *TP53*$^{-/-}$ control, 3,996 cells acquired and 2,180 cells retained; RPE *TP53*$^{-/-}$ 6-weeks post-WGD, 3,716 acquired and 3,475 cells retained; RPE *TP53*$^{-/-}$ 20-weeks post-WGD, 2,727 cells acquired and 1,976 cells retained; RPE *TP53*$^{-/-}$ 20-weeks post-WGD T1, 3,359 cells acquired and 2,851 cells retained; RPE *TP53*$^{-/-}$ 20-weeks post-WGD T2, 3,467 cells acquired and 2,721 cells retained; and RPE *TP53*$^{-/-}$ 20-weeks post-WGD, 3,790 cells acquired and 3,078 cells retained. On the intersection of genes across the dataset, 17,187 genes were found. The 17,187 genes were kept in all datasets and the merge function from the library was applied to merge the data. The dataset consisted of 17,187 genes and 21,055 cells in total. Cells having fewer than 200 and more than 10,000 genes expressed were removed from the analysis. Cells that had more than 8% of mitochondrial UMI genes expressed were also removed. The final dataset consisted of 17,187 genes and 16,281 cells. At this stage, the library depths were standardized using the NormalizeData function from Seurat, in which UMI counts for each cell were divided by the total UMI counts for that cell with a scaling factor set at 10,000. The expression matrix thus obtained was natural log-transformed. RunPCA function was run with default parameters and 2,000 highly variable genes (found using the FindVariableFeatures function). The RunUMAP function was run to obtain the uniform manifold approximation and projections, keeping default parameters and the number of PCA components, that is, ndim=1:12 was used.

scRNA-seq copy number changes using InferCNV

The InferCNV R package (v.1.1.0)[39] was used to call copy number changes. The UMI counts encompassing 17,187 genes and 21,055 cells were used as input to infer copy number changes. RPE control cells were considered as control samples. An inferCNV object was created using the CreateInfercnvObject object with raw UMI counts and hg19 genomic annotations as input. Run function parameters were set as follows: cutoff=0.1, cluster_by_groups=FALSE, denoise=TRUE, tumour_subcluster_partition_method = “qnorm”, HMM = TRUE, HMM_type = “i6”, analysis_mode = “subcluster”, HMM_report_by

="subcluster". The residuals matrix generated from InferCNV containing the copy number status was used to perform hierarchical clustering using the fastcluster (v.1.2.3) Python package[84]. The Elbow method was used to determine the optimal number of clusters[85], in which the distance between the clusters was plotted against each threshold. Matplotlib (v.3.4.2) library[86] was used for data visualization.

Differential expression analysis using scRNA-seq data

Differentially expressed genes between RPE *TP53*$^{-/-}$ control and 20-week post-WGD tumours were detected using MAST[87] from the Seurat package. All the genes with an adjusted *P* value below 0.001 and absolute $\log_2$(FC) greater than 0.3 were considered as differentially expressed.

Transcriptional regulator scores were then computed following the SCENIC workflow[88,89], using the pyscenic package v.0.11.2 as follows: the gene regulatory network was generated using the grn command, then the regulons (transcription factors and their target genes) were identified with using the ctx command using the motif list motifs-v9-nr.hgnc-m0.001-o0.0.tbl (downloaded from cisTarget database: https://resources.aertslab.org/cistarget/). The regulons for each single cell were scored using the aucell command.

Enrichment of differentially expressed genes on CoREs

The percentage of upregulated and downregulated genes overlapping with activating and inactivating CoREs were computed separately and used as background the complete set of genes. Effect sizes were then computed as $\log_2$(FC) for the four possible scenarios \({\sigma }_{{\rm{Up}}}^{{\rm{Activating}}},{\sigma }_{{\rm{Down}}}^{{\rm{Activating}}},\) \({\sigma }_{{\rm{Up}}}^{{\rm{Inactivating}}}\,{\rm{and}}\,{\sigma }_{{\rm{Down}}}^{{\rm{Inactivating}}}\) where, for example,

$${\sigma }_{{\rm{Up}}}^{{\rm{Activating}}}={\log }_{2}\frac{ \% \, {\rm{upregulated}}\,{\rm{genes}}\,{\rm{overlapping}}\,{\rm{an}}\,

{\rm{activating}}\,{\rm{CoRE}}}{ \% \,{\rm{all}}\,{\rm{genes}}\,
{\rm{overlapping}}\,{\rm{an}}\,{\rm{activating}}\,{\rm{CoRE}}}$$

and the other effect sizes were similarly computed.

Statistical significance was assessed by randomly selecting 100,000 times a set of genes with the same number of genes as the total number of differentially expressed genes and computing the percentage of those genes overlapping a CoRE. An empirical *P* value was calculated as the probability that a random set of genes has a percentage of overlapping genes higher than the real set of differentially expressed genes.

WGS

Library preparation

DNA was extracted from cells using a DNeasy Blood & Tissue kit (Qiagen, 69504) following the manufacturer's protocol. Library preparation was performed using TruSeq DNA PCR-Free (Illumina, 2001596) or TruSeq DNA Nano (Illumina, 2001596) kits. Libraries were sequenced on a NovaSeq 6000 system (Illumina) with a PE150 configuration.

Mapping and processing

Paired-end fastq files for each sample were aligned jointly to human_g1k_hs37d5 from the 1000 Genomes Phase 3 using bwa mem (v.0.7.17) and sorted with samtools (v.1.10)[90] using the sort command. Mismatches in read pairing were fixed using the fixmate command. Duplicate reads were identified using the genome analysis toolkit (GATK)[91,92] with the MarkDuplicatesSpark command. Base quality scores were corrected using BaseRecalibrator and ApplyBQSR from the GATK suite.

CNV calling

CNVs were identified using the Control-FREEC software[93] (v.11.6). CNVs for each sample were called with respect to the RPE *TP53*$^{-/-}$ control sample by taking as input the sample and the control bam files that were previously generated. The software was run using the following parameters:

ploidy = 2,3,4;

breakPointThreshold = .08;

intercept = 0;

window = 10000;

mateOrientation = FR;

sex = XX.

All other parameters were kept as default.

Statistical significance for each detected CNV was then computed using the assess_significance.R script provided with the Control-FREEC software.

Finding a consensus CNV set in RPE *TP53* $^{-/-}$ 20-weeks post-WGD tumours

The CNVs obtained using Control-FREEC having both a Kolmogorov–Smirnov *P* value and a Wilcoxcon *P* value below 0.01 were selected. Copy number changes of the same type (gain or loss) were merged on the basis of their overlap (±50 kb) across the three tumour samples using the merge function of bedtools (v.2.30.0)[64,65].

Mutation calling

SNVs were called using the Mutect2 (ref. [91]) algorithm from GATK. Blacklisted regions were derived from ref. [94] and excluded using the -XL option. F1R1 counts calculated in the pre-processing phase were provided with the option --f1r2-tar-gz. Germline variants were filtered out using the --germline-resource option with the gnomAD database (b37 version)[95]. Filtering of the mutations was performed using the FilterMutectCalls command setting --contamination-estimate to 0 and using the read orientation model computed at pre-processing with the --ob-priors option. VCF files were then converted to MAF using vcf2maf[96] and assembled into a single file. Only single-nucleotide polymorphism variants were considered in the analysis.

Mutations having the following FILTER values were removed: normal_artifact, germline, multiallelic and clustered_events. Then, all mutations having a gnomAD_AF value above 0.01 or were already reported in dbSNP (dbSNP_RS == ‘novel’) were removed. The VAF in the tumour

sample and in the control sample, respectively, were calculated as t_vaf = t_alt_count/t_depth and n_vaf = n_alt_count/n_depth. Mutations such that t_vaf < 2n_vaf were removed. For the remaining mutations, the overlap between samples was checked (n_samples). Mutations found in only one sample and having FILTER != 'PASS' or t_depth < 6 were removed. Conversely, if a mutation is found in only one sample but FILTER == 'PASS' and t_depth >= 6, it was kept.

Oncogenicity of the variants was assessed using OncoKB[97].

To study how many mutations were gained after WGD and to compare them to the ones accumulated in the same time frame without WGD induction, the number of mutations found in 6 weeks post-WGD and 6 weeks control samples were counted, with the following exclusion criteria: mutations already detected in the *TP53*$^{-/-}$ RPE control sample; mutations shared between 6-weeks post-WGD and 6-weeks control samples, as these mutations were most likely already present in the *TP53*$^{-/-}$ RPE control sample and went undetected; mutations shared between 6-weeks control and WGD samples for the same reason.

ChIP–seq and analysis

ChIP–seq

ChIP was performed using a SimpleChIP Enzymatic Chromatin IP kit (Cell Signaling Technology, 91820S) following the manufacturer's protocol. In brief, 2–4 million cells per condition were fixed in 1% formaldehyde for 10 min at room temperature. The reaction was quenched with 1× glycine and the cell pellet was washed with PBS. Cells were lysed, nuclei were digested with 1,000 U micrococcal nuclease for 20 min at 37 °C and briefly sonicated at 80 V peak incidence power, 200 cycles per burst and 5% duty factor for 90 s on an E220 focused-ultrasonicator (Covaris) to disrupt the nuclear membrane. Next 1–5 μg digested chromatin, with or without 0.1–0.5 μg digested mouse chromatin for spike-in normalization, was incubated at 4 °C overnight with one of the following antibodies, at the recommended dilutions: anti-acetyl-histone H3 (Lys27) (Cell Signaling Technology, 8173), anti-trimethyl-histone H3 (Lys9) (Cell Signaling Technology,

13969), anti-CTCF (active motif, 61311), anti-trimethyl-histone H3 (Lys4) (Cell Signaling Technology, 9751) or trimethyl-histone H3 (Lys27) (Cell Signaling Technology, 9733). Antibody-bound chromatin was precipitated using protein G magnetic beads and crosslink reversal was performed using NaCl and proteinase K. Resulting DNA was purified using spin columns. Library preparation for sequencing from the chromatin immunoprecipitated DNA was performed using a NEBNext Ultra II DNA Library Prep kit for Illumina (New England BioLabs, E7645) following the manufacturer's protocol. DNA fragments were processed for end repair, followed by stubby adaptor ligation. DNA was size-selected using AMPure XP beads (Beckman-Coulter, A63881). Adaptor-ligated DNA was PCR-amplified using indexed NEBNext Multiplex Oligos for Illumina (New England Biolabs, E7335). The resulting libraries were sequenced on a NextSeq 500 system (Illumina) in a PE37/38 configuration.

ChIP–seq data analysis

Fastq files were processed using the nfcore/chipseq pipeline (v.1.2.2)[98,99] (https://nf-co.re/chipseq/) with default parameters, aligning the reads to the hg19 genome. For each sample, the fold change against the input experiment was then computed using the bamComapre command from the deepTools package (v.3.5.1)[100] setting –scaleFactorsMethod readCount, --extendReads, --operation ratio and --binSize 100). MACS3 (ref. [101]) was used for peak calling. Peaks with a *q* value lower than 0.1 and a minimum fold change of 1.5 were retained. A consensus set of peaks for CTCF and H3K9me3 was then created from control and WGD samples from the RPE $TP53^{-/-}$ and CP-A $TP53^{-/-}$ cell lines separately by aggregating peaks that were closer than 10 kb for H3K9me3 and that directly overlapped for CTCF in the two samples. The maximum ChIP–seq signal (fold change over input) in control and WGD was associated for each consensus peak in the collection.

Spike-in normalization

Spike-in normalization on the CTCF signal was performed as previously described[102].

CTCF peaks at Hi-C boundaries

Insulation boundaries shared between control and WGD samples from RPE *TP53*$^{-/-}$ samples identified from Hi-C were associated to CTCF peaks by taking all the peaks lying in on the genomic bin of the boundary (±10 kb). Boundaries were then associated to a CTCF score representing the sum of all the CTCF peak signals at the boundary.

Comparing ChIP–seq signal differences to Hi-C compartment differences

Fold change signals computed at the previous step were scaled for each 10/50 bp bin (having value x) computing the $\log_2(x + 1)$ value. Scaled values were then binned at 50 kb resolution, taking the average signal in the bin. Finally, the 50 kb binned signal was normalized by dividing each bin by the chromosome median value. For each RPE *TP53*$^{-/-}$ 20-weeks post-WGD tumour Hi-C sample, the binned Calder ranks at 50 kb were matched to the binned ChIP–seq values for each histone modification. Finally, differences in Calder rank (Δrank) were compared with differences in ChIP–seq signal (Δhm_1, Δhm_2, …) for each tumour against the RPE *TP53*$^{-/-}$ control sample.

Correlation between CoRE regions and differences in ChIP–seq signal

To study the relationship between subcompartment repositioning events and epigenetic changes in RPE *TP53*$^{-/-}$ post-WGD tumours, the average ChIP–seq signal difference for each histone mark was assigned to each CoRE (average(Δhm_1), average(Δhm_2), …), where $\Delta {{\rm{hm}}}_{1}={{\rm{hm}}}_{1}^{{\rm{Tumour}}}-{{\rm{hm}}}_{1}^{{\rm{Control}}}$ for each 50 kb bin in the CoRE region. The Spearman correlation was computed between subcompartment repositioning scores and the average epigenetic differences of the CoREs. To estimate the significance of the observed correlations, an empirical P value was computed by randomly sampling a number of regions of the same size of the observed CoREs across the genome 1,000 times and

recomputing the correlation value. A *P* value was obtained, corresponding to the number of times the expected correlation was greater or equal in absolute value to the one observed, divided by the total number of random trials.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All raw data have been deposited into the NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE222391. Processed Hi-C data together with compartment domain calls by Calder are available at https://doi.org/10.5281/zenodo.7351767. ChIP–seq data, scRNA-seq matrices and copy number profiles are available at https://doi.org/10.5281/zenodo.7351776. Source data are provided with this paper.

Code availability

Analysis code regarding LCS and the identification of CoREs is available at https://github.com/CSOgroup/WGD.

References

1. Fujiwara, T. et al. Cytokinesis failure generating tetraploids promotes tumorigenesis in p53-null cells. *Nature* **437**, 1043–1047 (2005).

2. Olaharski, A. J. et al. Tetraploidy and chromosomal instability are early events during cervical carcinogenesis. *Carcinogenesis* **27**, 337–343 (2006).

3. Bielski, C. M. et al. Genome doubling shapes the evolution and prognosis of advanced cancers. *Nat. Genet.* **50**, 1189 (2018).

4. Davoli, T. & de Lange, T. The causes and consequences of polyploidy in normal development and cancer. *Annu. Rev. Cell Dev. Biol.* **27**, 585–610 (2011).

5. Dewhurst, S. M. et al. Tolerance of whole-genome doubling propagates chromosomal instability and accelerates cancer genome evolution. *Cancer Discov.* **4**, 175–185 (2014).

6. Kuznetsova, A. Y. et al. Chromosomal instability, tolerance of mitotic errors and multidrug resistance are promoted by tetraploidization in human cells. *Cell Cycle* **14**, 2810–2820 (2015).

7. Gemble, S. et al. Genetic instability from a single S phase after whole-genome duplication. *Nature* **604**, 146–151 (2022).

8. Duelli, D. M. et al. A virus causes cancer by inducing massive chromosomal instability through cell fusion. *Curr. Biol.* **17**, 431–437 (2007).

9. Galipeau, P. C. et al. 17p (p53) allelic losses, 4N (G2/tetraploid) populations, and progression to aneuploidy in Barrett's esophagus. *Proc. Natl Acad. Sci. USA* **93**, 7081–7084 (1996).

10. Barrett, M. T. et al. Molecular phenotype of spontaneously arising 4N (G2-tetraploid) intermediates of neoplastic progression in Barrett's esophagus. *Cancer Res.* **63**, 4211–4217 (2003).

11. Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).

12. Rao, S. S. P. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).

13. Bonev, B. & Cavalli, G. Organization and function of the 3D genome. *Nat. Rev. Genet.* **17**, 661–678 (2016).

14. Wang, L. et al. Histone modifications regulate chromatin compartmentalization by contributing to a phase separation mechanism. *Mol. Cell* **76**, 646–659.e6 (2019).

15. Sanborn, A. L. et al. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl Acad. Sci. USA* **112**, E6456–E6465 (2015).

16. Nuebler, J., Fudenberg, G., Imakaev, M., Abdennur, N. & Mirny, L. A. Chromatin organization by an interplay of loop extrusion and compartmental segregation. *Proc. Natl Acad. Sci. USA* **115**, E6697–E6706 (2018).

17. Dixon, J. R. et al. Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**, 331–336 (2015).

18. Flavahan, W. A. et al. Insulator dysfunction and oncogene activation in *IDH* mutant gliomas. *Nature* **529**, 110–114 (2016).

19. Flavahan, W. A. et al. Altered chromosomal topology drives oncogenic programs in SDH-deficient GISTs. *Nature* **575**, 229–233 (2019).

20. Weischenfeldt, J. et al. Pan-cancer analysis of somatic copy-number alterations implicates *IRS4* and *IGF2* in enhancer hijacking. *Nat. Genet.* **49**, 65–74 (2017).

21. Zimmerman, M. W. et al. *MYC* drives a subset of high-risk pediatric neuroblastomas and is activated through mechanisms including enhancer hijacking and focal enhancer amplification. *Cancer Discov.* **8**, 320–335 (2018).

22. Sungalee, S. et al. Histone acetylation dynamics modulates chromatin conformation and allele-specific interactions at oncogenic loci. *Nat. Genet.* **53**, 650–662 (2021).

23. Donaldson-Collier, M. C. et al. EZH2 oncogenic mutations drive epigenetic, transcriptional, and structural changes within chromatin domains. *Nat. Genet.* **51**, 517 (2019).

24. Rosencrance, C. D. et al. Chromatin hyperacetylation impacts chromosome folding by forming a nuclear subcompartment. *Mol. Cell* **78**, 112–126.e12 (2020).

25. Johnstone, S. E. et al. Large-scale topological changes restrain malignant progression in colorectal cancer. *Cell* **182**, 1474–1489.e23 (2020).

26. Stachler, M. D. et al. Paired exome analysis of Barrett's esophagus and adenocarcinoma. *Nat. Genet.* **47**, 1047–1055 (2015).

27. Drainas, A. P. et al. Genome-wide screens implicate loss of cullin ring ligase 3 in persistent proliferation and genome instability in *TP53*-deficient cells. *Cell Rep.* **31**, 107465 (2020).

28. Law, J. C., Ritke, M. K., Yalowich, J. C., Leder, G. H. & Ferrell, R. E. Mutational inactivation of the *p53* gene in the human erythroid leukemic K562 cell line. *Leuk. Res.* **17**, 1045–1050 (1993).

29. Liu, Y. et al. Systematic inference and comparison of multi-scale chromatin sub-compartments connects spatial organization to cell phenotypes. *Nat. Commun.* **12**, 2439 (2021).

30. Crane, E. et al. Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature* **523**, 240–244 (2015).

31. Bakhoum, S. F. et al. Chromosomal instability drives metastasis through a cytosolic DNA response. *Nature* **553**, 467–472 (2018).

32. Nora, E. P. et al. Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell* **169**, 930–944.e22 (2017).

33. Rao, S. S. P. et al. Cohesin loss eliminates all loop domains. *Cell* **171**, 305–320.e24 (2017).

34. Ganem, N. J. et al. Cytokinesis failure triggers Hippo tumor suppressor pathway activation. *Cell* **158**, 833–848 (2014).

35. Chadwick, B. P. Chromatin of the Barr body: histone and non-histone proteins associated with or excluded from the inactive X chromosome. *Hum. Mol. Genet.* **12**, 2167–2178 (2003).

36. Baslan, T. et al. Ordered and deterministic cancer genome evolution after p53 loss. *Nature* **608**, 795–802 (2022).

37. Burd, C. E. et al. Mutation-specific RAS oncogenicity explains NRAS codon 61 selection in melanoma. *Cancer Discov.* **4**, 1418–1429 (2014).

38. Shoshani, O. et al. Transient genomic instability drives tumorigenesis through accelerated clonal evolution. *Genes Dev.* **35**, 1093–1108 (2021).

39. Patel, A. P. et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344**, 1396–1401 (2014).

40. Mardin, B. R. et al. A cell-based model system links chromothripsis with hyperploidy. *Mol. Syst. Biol.* **11**, 828 (2015).

41. Gisselsson, D. et al. Chromosomal breakage-fusion-bridge events cause genetic intratumor heterogeneity. *Proc. Natl Acad. Sci. USA* **97**, 5357–5362 (2000).

42. Reymond, N. et al. Cdc42 promotes transendothelial migration of cancer cells through β1 integrin. *J. Cell Biol.* **199**, 653–668 (2012).

43. Eferl, R. & Wagner, E. F. AP-1: a double-edged sword in tumorigenesis. *Nat. Rev. Cancer* **3**, 859–868 (2003).

44. Laoukili, J. et al. FoxM1 is required for execution of the mitotic programme and chromosome stability. *Nat. Cell Biol.* **7**, 126–136 (2005).

45. Anastas, J. N. & Moon, R. T. WNT signalling pathways as therapeutic targets in cancer. *Nat. Rev. Cancer* **13**, 11–26 (2013).

46. Silver, D. P. & Livingston, D. M. Mechanisms of BRCA1 tumor suppression. *Cancer Discov.* **2**, 679–684 (2012).

47. Lee, M.-N. et al. Epigenetic inactivation of the chromosomal stability control genes *BRCA1, BRCA2*, and *XRCC5* in non–small cell lung cancer. *Clin. Cancer Res.* **13**, 832–838 (2007).

48. Asbaghi, Y., Thompson, L. L., Lichtensztejn, Z. & McManus, K. J. *KIF11* silencing and inhibition induces chromosome instability that may contribute to cancer. *Genes Chromosomes Cancer* **56**, 668–680 (2017).

49. McKenna, A. et al. Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science* **353**, aaf7907 (2016).

50. Biddy, B. A. et al. Single-cell mapping of lineage and identity in direct reprogramming. *Nature* **564**, 219–224 (2018).

51. Wang, R. W., MacDuffie, E. & Santaguida, S. Generation and isolation of cell cycle-arrested cells with complex karyotypes. *J. Vis. Exp.* https://doi.org/10.3791/57215 (2018).

52. Santaguida, S., Tighe, A., D'Alise, A. M., Taylor, S. S. & Musacchio, A. Dissecting the role of MPS1 in chromosome biorientation and the spindle checkpoint through the small molecule inhibitor reversine. *J. Cell Biol.* **190**, 73–87 (2010).

53. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).

54. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at https://doi.org/10.48550/arXiv.1303.3997 (2013).

55. Durand, N. C. et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98 (2016).

56. Knight, P. A. & Ruiz, D. A fast algorithm for matrix balancing. *IMA J. Numer. Anal.* **33**, 1029–1047 (2013).

57. Yang, T. et al. HiCRep: assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient. *Genome Res.* **27**, 1939–1949 (2017).

58. Lin, D., Sanders, J. & Noble, W. S. HiCRep.py: fast comparison of Hi-C contact matrices in Python. *Bioinformatics* **37**, 2996–2997 (2021).

59. Pfitzner, D., Leibbrandt, R. & Powers, D. Characterization and evaluation of similarity measures for pairs of clusterings. *Knowl. Inf. Syst.* **19**, 361–394 (2009).

60. Kruse, K., Hug, C. B. & Vaquerizas, J. M. FAN-C: a feature-rich framework for the analysis and visualisation of chromosome conformation capture data. *Genome Biol.* **21**, 303 (2020).

61. Imakaev, M. et al. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat. Methods* **9**, 999–1003 (2012).

62. Abdennur, N. & Mirny, L. A. Cooler: scalable storage for Hi-C data and other genomically labeled arrays. *Bioinformatics* **36**, 311–316 (2020).

63. Vilarrasa-Blasi, R. et al. Dynamics of genome architecture and chromatin function during human B cell differentiation and neoplastic transformation. *Nat. Commun.* **12**, 651 (2021).

64. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

65. Dale, R. K., Pedersen, B. S. & Quinlan, A. R. Pybedtools: a flexible Python library for manipulating genomic datasets and annotations. *Bioinformatics* **27**, 3423–3424 (2011).

66. Olshen, A. B., Venkatraman, E. S., Lucito, R. & Wigler, M. Circular binary segmentation for the analysis of array-based DNA copy number

data. *Biostatistics* **5**, 557–572 (2004).

67. Garrison, E. & Marth, G. Haplotype-based variant detection from short-read sequencing. Preprint at https://doi.org/10.48550/arXiv.1207.3907 (2012).

68. Bansal, V. Integrating read-based and population-based phasing for dense and accurate haplotyping of individual genomes. *Bioinformatics* **35**, i242–i248 (2019).

69. Delaneau, O., Zagury, J.-F. & Marchini, J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat. Methods* **10**, 5–6 (2013).

70. Edge, P., Bafna, V. & Bansal, V. HapCUT2: robust and accurate haplotype assembly for diverse sequencing technologies. *Genome Res.* **27**, 801–812 (2017).

71. Carty, M. et al. An integrated model for detecting significant chromatin interactions from high-resolution Hi-C data. *Nat. Commun.* **8**, 15454 (2017).

72. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).

73. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

74. Chen, E. Y. et al. Enrichr: interactive and collaborative *HTML5* gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128 (2013).

75. Nagano, T. et al. Single-cell Hi-C for genome-wide detection of chromatin interactions that occur simultaneously in a single cell. *Nat. Protoc.* **10**, 1986–2003 (2015).

76. Bonzon, D. et al. Impedance-based single-cell pipetting. *SLAS Technol. Transl. Life Sci. Innov.* **25**, 222–233 (2020).

77. Wolff, J. et al. Galaxy HiCExplorer: a web server for reproducible Hi-C data analysis, quality control and visualization. *Nucleic Acids Res.* **46**, W11–W16 (2018).

78. Nagano, T. et al. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* **502**, 59–64 (2013).

79. Rousseeuw, P. J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **20**, 53–65 (1987).

80. Lloyd, S. Least squares quantization in PCM. *IEEE Trans. Inf. Theory* **28**, 129–137 (1982).

81. Rand, W. M. Objective criteria for the evaluation of clustering methods. *J. Am. Stat. Assoc.* **66**, 846–850 (1971).

82. Zheng, G. X. Y. et al. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).

83. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e21 (2019).

84. Müllner, D. fastcluster: fast hierarchical, agglomerative clustering routines for R and Python. *J. Stat. Softw.* **53**, 1–18 (2013).

85. Thorndike, R. L. Who belongs in the family? *Psychometrika* **18**, 267–276 (1953).

86. Hunter, J. D. Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).

87. Finak, G. et al. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* **16**, 278 (2015).

88. Aibar, S. et al. SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).

89. Van de Sande, B. et al. A scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nat. Protoc.* **15**, 2247–2276 (2020).

90. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

91. Auwera, G. A. et al. From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Curr. Protoc. Bioinformatics* **43**, 11.10.1–11.10.33 (2013).

92. McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

93. Boeva, V. et al. Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics* **28**, 423–425 (2012).

94. Amemiya, H. M., Kundaje, A. & Boyle, A. P. The ENCODE Blacklist: identification of problematic regions of the genome. *Sci Rep.* **9**, 9354 (2019).

95. Karczewski, K. J. et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).

96. Kandoth, C. et al. Mskcc/Vcf2Maf: Vcf2Maf v1.6.16. *Zenodo* https://doi.org/10.5281/zenodo.1185418 (2018).

97. Chakravarty, D. et al. OncoKB: a precision oncology knowledge base. *JCO Precis. Oncol.* https://doi.org/10.1200/PO.17.00011 (2017).

98. Patel, H. et al. nf-core/chipseq: nf-core/chipseq v1.2.2—Rusty Mole. *Zenodo* https://doi.org/10.5281/zenodo.4711243 (2021).

99. Ewels, P. A. et al. The nf-core framework for community-curated bioinformatics pipelines. *Nat. Biotechnol.* **38**, 276–278 (2020).

100. Ramírez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).

101. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).

102. Orlando, D. A. et al. Quantitative ChIP-Seq normalization reveals global modulation of the epigenome. *Cell Rep.* **9**, 1163–1170 (2014).

Acknowledgements

We thank J. Korbel for sharing the RPE cells and RPE *TP53*$^{-/-}$ cells used in this study; and the staff at the EPFL facilities (the Flow Cytometry Core Facility, the Bioimaging and Optics Platform, the Center of PhenoGenomics), in particular the team at the Gene Expression Core Facility led by B. Mangeat. This work was supported by the EPFL (to E.O.) and UNIL (to G.C.) internal funds, the Swiss National Science Foundation (to E.O. and G.C.) and the Swiss Cancer League (to E.O. and G.C.).

Funding

Open access funding provided by EPFL Lausanne.

Author information

Author notes

1. These authors contributed equally: Ruxandra A. Lambuta, Luca Nanni

Authors and Affiliations

1. Swiss Institute for Experimental Cancer Research (ISREC), School of Life Sciences, EPFL, Écublens, Switzerland

   Ruxandra A. Lambuta, Juan Diaz-Miyar, Natalya Katanayeva & Elisa Oricchio

2. Swiss Cancer Center Leman, Lausanne, Switzerland

   Ruxandra A. Lambuta, Luca Nanni, Yuanlong Liu, Juan Diaz-Miyar, Arvind Iyer, Daniele Tavernari, Natalya Katanayeva, Giovanni Ciriello & Elisa Oricchio

3. Department of Computational Biology, University of Lausanne (UNIL), Lausanne, Switzerland

   Luca Nanni, Yuanlong Liu, Arvind Iyer, Daniele Tavernari & Giovanni Ciriello

4. Swiss Institute of Bioinformatics (SIB), Lausanne, Switzerland

   Luca Nanni, Yuanlong Liu, Arvind Iyer, Daniele Tavernari & Giovanni Ciriello

Contributions

R.A.L. designed and performed all the experiments. L.N. designed and developed new methods to analyse the data and performed all the computational analyses. Y.L. contributed to Hi-C data analyses. J.D.-M. contributed to the analyses of genomic data. A.I. and D.T. contributed to scRNA-seq data analyses. N.K. contributed to sample preparation. G.C. and E.O. designed and supervised the study. R.A.L., L.N., G.C. and E.O. wrote the manuscript with input from all the authors.

Corresponding authors

Correspondence to Giovanni Ciriello or Elisa Oricchio.

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Whole-genome doubling induction.

a, Immunoblot presenting TP53 and beta-actin protein levels in RPE and CP-A, *TP53* WT and *TP53*$^{-/-}$ cell lines. Samples treated with the TP53-pathway inducer doxorubicin (DXR) were included. (n = 1 experiment) **b**, Example of PI-based cell cycle staining histograms for control and WGD populations of CP-A$^{TP53-/-}$ (clone 19 and clone 3) and K562 cells. **c**, Percentage of cells of different ploidies for CP-A$^{TP53-/-}$ C3 control (n = 5) and WGD (n = 5); RPE$^{TP53-/-}$ control (n = 10), WGD (n = 11), WGD CDK1i+DCB (n = 3), and CIN-only (n = 2); K562 control (n = 3) and WGD (n = 4); and CP-A$^{TP53-/-}$ clone 19 control (n = 6) and WGD (n = 6) populations. Data are mean ± s.d.; all data points are shown; n indicates the number of independent experiments. **d**, Immunofluorescence images of RPE$^{TP53-/-}$ cells showing transition from mononucleated diploids to mononucleated tetraploids via binucleated intermediates during WGD induction. Nuclei are shown in blue (DAPI) and cytoskeleton (alpha-tubulin) in red (n = 100 cells from 2 WGD inductions per condition were considered). **e**, Percentage of binucleated cells in RPE$^{TP53-/-}$ control, at WGD-induction treatment removal, and 24 and 48 h (WGD point) post-treatment removal populations. **f**, Example of PI-based cell cycle staining histograms for RPE$^{TP53-/-}$ control, WGD induced via nocodazole (Noc) +

DCB, WGD induced via CDK1i+DCB treatment before and after cell sorting, and CIN-only populations. **g**, Karyotyping of CP-A$^{TP53-/-}$ clone 19 control and WGD cells. Example images of metaphase spreads and quantification of chromosomes per cell are shown. **h**, Relative nucleus area measured for control and WGD CP-A$^{TP53-/-}$ (clone 19 and clone 3), and RPE$^{TP53-/-}$ cells (n = 100, from 2 independent WGD inductions).

Source data

Extended Data Fig. 2 Chromatin topological similarities between control and WGD.

a-e, Comparisons of Hi-C contact maps between control diploid cells and replicate (R) or WGD cells of the indicated cell line, or between control diploid cells of the indicated cell line and GM12878 cells. **a**, Stratum correlation coefficient of matching intra-chromosomal maps (n = 23 per boxplot), **b**, inter-chromosomal Hi-C contact correlation of matching inter-chromosomal maps (n = 253), **c**, Spearman's correlation of compartment domain ranks computed by Calder of matching intra-chromosomal maps (n = 23), **d**, measure of concordance of TADs for each matching intra-chromosomal maps (n = 23), **e**, Spearman's correlation of insulation scores of matching intra-chromosomal maps (n = 23). Boxplots: central line is the median, bounding box corresponds to the 25th–75th percentiles, and the whiskers extend up to 1.5 times the interquartile range. **f**, Examples of intrachromosomal loss of chromatin segregation in RPE$^{TP53-/-}$ (top), CP-A$^{TP53-/-}$ C3 (centre) and K562 (bottom) models. Intrachromosomal observed/expected maps in Control and WGD samples are shown on the right in log2 scale, with Control in the lower and WGD (left) or Observed/Expected (right) in the upper triangular matrices. **g**, Relationship between observed/expected intra-chromosomal values in control vs. WGD samples. Each dot represents an inter-chromosomal Hi-C interaction. Red dots denote higher, whereas blue dots lower interaction signals in WGD compared to control.

Source data

Extended Data Fig. 3 Chromatin loss of segregation and loss of boundary insulation at WGD.

a, Balanced inter-chromosomal contacts in all control, WGD, and CIN-only samples. Colour scales are centred in 1/23. Two clusters of interactions are highlighted: long chromosomes (chr1 to chr14 and chrX), and short chromosomes (chr15 to chr22). **b**, Ratios of inter-chromosomal contact enrichments (observed vs. expected) between WGD and control samples for CP-A$^{TP53-/-}$ clone 19 and RPE$^{TP53-/-}$ (Noc+DCB), an independent WGD induction (WGD2). Chromosomes were sorted by length. Long chromosomes (chr1 to chr14 and chrX) and short chromosomes (chr15 to chr22) are highlighted. *P*-values are calculated by two-tailed Wilcoxon test. **c**, Heatmap of ratios of genomic bins belonging to the indicated sub-compartments that gain vs. lose contacts for CP-A$^{TP53-/-}$ clone 19 and RPE$^{TP53-/-}$ (Noc+DCB), an independent WGD induction (WGD2). *P*-values were calculated by two-tailed Wilcoxon test. **d**, Boundary insulation scores in Control (x-axis) and WGD or CIN-only (y-axis) samples. Only boundaries shared between the two conditions are considered. *P*-values are calculated by two-tailed Wilcoxon test. **e**, Ratios of inter-chromosomal contact enrichments (left), genomic bins belonging to the indicated sub-compartments that gain vs. lose contacts (middle), and boundary insulation scores (right) between samples with different sequencing coverage to account for coverage per genome size discrepancies between the control and WGD conditions. Conditions considered were RPE$^{TP53-/-}$ WGD (219 million reads) vs control (108 million reads), and RPE$^{TP53-/-}$ control (108 million reads) versus control (221 million reads). All *P*-values are calculated by two-tailed Wilcoxon test. **f**, Fraction of interactions between homologous copies of each chromosome in RPE$^{TP53-/-}$ control and WGD cells.

Source data

Extended Data Fig. 4 Effect of WGD on the transcriptome and on CTCF and H3K9me3-mediated chromatin compaction.

a, Volcano plot showing the fold change and its associated adjusted *P*-value calculated by MAST for each differentially expressed genes between WGD condition and control in $RPE^{TP53-/-}$ cells. Significantly upregulated genes are shown in yellow, whereas significantly downregulated genes in blue. **b, c**, Gene set enrichment analysis for the significantly upregulated **(b)** and significantly downregulated **(c)** genes in $RPE^{TP53-/-}$ WGD versus control cells. Databases utilized for the analysis for each gene set are indicated. Adjusted *p*-value calculated by hypergeometric test. **d**, Differentially expressed genes enrichment at boundaries that gain or lose insulation post-WGD. Only boundaries having an insulation score in Control below −0.1 were considered. The enrichment is calculated using as background the complete set of genes. **e**, Distribution of compartment rank differences between Control and WGD $RPE^{TP53-/-}$ cells estimated by Calder (Δcompartment rank: negative values indicate a shift towards less active compartment, positive values indicate a shift towards a more active compartment) in bins overlapping gene promoters. Distributions relative to genes not changing expression (grey), downregulated genes (blue) and upregulated genes (orange) are shown. **f**, Quantification of indicated proteins in diploid control and WGD CP-$A^{TP53-/-}$ (clones 3 and 19) and $RPE^{TP53-/-}$ cells. Data is mean ± s.d. **g**, Number of CTCF peaks identified in $RPE^{TP53-/-}$ control and WGD samples and their overlap. **h**, ChIP-seq signal of CTCF shared peaks between control and WGD conditions in CP-$A^{TP53-/-}$ clone 3 cells. Dots are coloured by point density in log10 scale, and the regression line (red) is calculated using locally weighted scatterplot smoothing (LOWESS). **i**, Aggregated CTCF peak signals in $RPE^{TP53-/-}$ control cells at boundaries that gain (*red*) or lose (*blue*) insulation in WGD cells. Top-300 boundaries (*left*, gain n = 62, lose n = 81) and boundaries having an absolute change of insulation greater than 0.1 (*right*, gain n = 340, lose n = 101) are considered. *p*-values are computed with a two-sided Mann-Whitney U test. Boxplots: central line is the median, bounding box corresponds to the 25th–75th percentiles, and the whiskers extend up to 1.5 times the interquartile range **j**, H3K9me3 enrichment across chromatin sub-compartments and changes in H3K9me3 signal, based on ChIP-Seq data, between WGD and control conditions in CP-$A^{TP53-/-}$ clone 3 (above) and $RPE^{TP53-/-}$ (below) cells. **k**, Schematic representation of WGD induction in *TP53* wild type CP-A cells and in CP-$A^{TP53-/-}$ treated with 0.5 µM of the

CDK4/6 inhibitor Palbociclib. **l**, Example of PI-based cell cycle staining histograms for control and WGD populations of CP-A cells. **m**, Percentage of cells of distinct ploidies as of CP-A control (n = 8) and WGD (n = 8) populations. Data are mean ± s.d.; n indicates the number of independent experiments. **n**, Karyotyping of CP-A control and WGD cells. Example images of metaphase spreads and quantification of chromosomes per cell are shown. **o**, Immunofluorescence images of RPE cells, Control and WGD, illustrating the maintenance of binucleated cells post-WGD. Nuclei are shown in blue (DAPI) and cytoskeleton (alpha-tubulin) in red (n = 100 cells from 2 WGD inductions per condition were tested). **p**, Quantification of indicated proteins in diploid control and WGD CP-A cells and in WGD CP-A$^{TP53-/-}$ cells treated with CDK4/6 inhibitor. Data is mean ± s.d. **q**, Ratios of inter-chromosomal contact enrichments (observed vs. expected) between CP-A and CP-A$^{TP53-/-}$ treated with CDK4/6i WGD and control samples. Chromosomes were sorted by length. Long chromosomes (chr1 to chr14 and chrX) and short chromosomes (chr15 to chr22) are highlighted. *p*-values are calculated by two-tailed Wilcoxon test.

Source data

Extended Data Fig. 5 *TP53* loss is not the cause of LCS.

a, Ratios of inter-chromosomal contact enrichments (observed vs. expected) between *TP53*$^{-/-}$ and *TP53* WT samples for CP-A and RPE cells. Chromosomes were sorted by length. Long chromosomes (chr1 to chr14 and chrX) and short chromosomes (chr15 to chr22) are highlighted. *p*-values are calculated by two-tailed Wilcoxon test. **b**, Heatmap of ratios of genomic bins belonging to the indicated sub-compartments that gain vs. lose contacts between *TP53*$^{-/-}$ and TP53 WT samples for CP-A and RPE cells. **c**, Boundary insulation scores in *TP53* WT and *TP53*$^{-/-}$ for CP-A and RPE samples. Only boundaries shared between the two conditions are considered. **b-c** *p*-values calculated by two-tailed Wilcoxon test. **d**, Representative image of immunoblot (above) and quantification (below) of CTCF protein expression in *TP53* WT and *TP53*$^{-/-}$ CP-A (left) and RPE (right) cells (n = 1 experiment). **e**, *CTCF* mRNA expression levels in RPE and CP-A *TP53* WT cell lines after treatment with 3 µM doxorubicin (n = 2 independent experiments).

Source data

Extended Data Fig. 6 scHi-C recovers loss of chromosome segregation and chromosomal instability features.

a, Chromosome interactions in pseudo-bulk Hi-C data derived from the aggregation of single-cell Hi-C data for RPE$^{TP53-/-}$ control and WGD, and Log2 fold-changes between WGD and Control (right) conditions. For individual conditions, interactions are shown as the Pearson correlation of the chromosome level Observed/Expected values for each pair of chromosomes. For WGD/Control condition, control and WGD correlation values were increased by 1 before computing the ratio. **b**, Fraction of altered genome in RPE$^{TP53-/-}$ control, WGD with no LCS, and WGD/LCS single cell groups. *p*-value calculated by two-tailed Wilcoxon test. Boxplots: central line is the median, bounding box corresponds to the 25th–75th percentiles, and the whiskers extend up to 1.5 times the interquartile range. **c**, Relative copy number changes in each of the single cells queried by scHi-C. Cells are sorted by fraction of altered genome.

Source data

Extended Data Fig. 7 Chromosomal instability and rapid aneuploidization post-WGD.

Changes in ploidy short term post-WGD (24 and 48 h) in CP-A$^{TP53-/-}$ (clone 19 and clone 3), and RPE$^{TP53-/-}$ as shown by **a,b**, PI-based cell cycle staining **(a)**, quantification of percentage of cells of distinct ploidies (2N, 4N, >4N) **(b)** and **c**, Quantification of chromosomes per cell via karyotyping. Data are mean ± s.d.; Number of independent experiments is indicated. Violin plots: dashed line is median **d**, Examples of chromosomal instability markers (telomere fusion and chromosome breakages) in CP-A$^{TP53-/-}$ clone 19 cells 24 h post-WGD. n = 3 out of 30 total karyotypes presented such markers. Red arrows indicate affected chromosomes. **e**, Examples of bipolar and multipolar division in RPE$^{TP53-/-}$ cells (Control and 24 h post-WGD). n = 5 dividing cells were detected for each phenotype

out of ~100 total cells. Nuclei are stained in blue (DAPI), cytoskeleton (alpha-tubulin) in red, and centrosomes (pericentrin) in green.

Source data

Extended Data Fig. 8 Tumour progression, ploidy, copy number changes, and gene expression changes post-WGD.

a, Evolution of the tumour volumes from the time of subcutaneous injection in NSG mice. Tumours were derived from RPE$^{TP53-/-}$ Control (n = 3), 6 weeks post-WGD (n = 9), and 20 weeks post-WGD (n = 3) cells. **b**, Mutational burden of RPE$^{TP53-/-}$ *in vitro* and *in vivo* samples defined as the number of SNPs compared to control. Control cells maintained in cell culture for 6 weeks were included (6w control). Variant allele frequency is displayed for 6w control, WGD, and 6w-pWGD samples. **c**, Breakpoints associated with copy number variants across the RPE$^{TP53-/-}$ 20 weeks post-WGD tumour samples. Copy number profile is shown as coverage ratio between tumour samples and control. Red lines mark shared breakpoints, green lines mark sample-specific breakpoints. Breakpoints defining the main CNVs shared by all tumour samples are highlighted with dashed red lines. **d**, Copy number alterations determined via WGS in RPE$^{TP53-/-}$ Control and post-WGD cells (in vitro and in vivo) derived from distinct WGD induction experiments (E2, E3, E4) **e**, Fraction of altered genome calculated for each single cell, using copy number changes derived by the InferCNV pipeline applied to the single-cell RNA-seq data for the RPE$^{TP53-/-}$ post-WGD samples (in vivo and in vitro). Boxplots: central line is the median, bounding box corresponds to the 25th–75th percentiles, and the whiskers extend up to 1.5 times the interquartile range. **f**, Estimated chromosomal copy number alterations of chromosome 13q from WGS in RPE$^{TP53-/-}$ 20-weeks post-WGD tumour 2. **g**, Volcano plot showing differential expression of significant genes in RPE$^{TP53-/-}$ control versus 20 weeks post-WGD derived Tumour 2 sample. Log2 fold-change of gene expression and adjusted *p*-value were calculated with the MAST algorithm. Genes overlapping a CNV are highlighted in blue (losses) and red (gains). **h**, Highest scoring transcription regulators of differentially expressed genes in the 20 weeks post-WGD derived Tumour 2 identified using the SCENIC

workflow. Expression changes of the transcription factors' targets are presented on the x-axis, while on the y-axis shows the SCENIC score for each regulon.

Source data

Extended Data Fig. 9 Loss of chromatin segregation in post-WGD tumours.

a, Ratios of inter-chromosomal contact enrichments (observed vs. expected) between RPE$^{TP53-/-}$ post-WGD tumours and control. Chromosomes were sorted by length. Long chromosomes (chr1 to chr14 and chrX) and short chromosomes (chr15 to chr22) are highlighted. **b**, Heatmap of ratios of genomic bins belonging to the indicated sub-compartments that gain vs. lose contacts between RPE$^{TP53-/-}$ post-WGD tumours and control. **c**, Boundary insulation scores in RPE$^{TP53-/-}$ control and post-WGD tumours. Only boundaries shared between the two conditions are considered. **a**–**c** *p*-values are calculated by two-tailed Wilcoxon test.

Source data

Extended Data Fig. 10 Properties of compartment repositioning events in post-WGD tumours and their consistency across different change-point detection algorithms.

a, Distribution of compartment rank differences estimated by Calder (Δcompartment rank: negative values indicate a shift towards less active compartment, positive values indicate a shift towards a more active compartment) between RPE$^{TP53-/-}$ control replicates (grey fill), WGD vs control (black line), and 20 weeks post-WGD tumours vs control (red line). **b**, Distribution of sub-compartment differences estimated by Calder (Δsub-compartment) in WGD and 20 weeks post-WGD tumours, compared to control. **c**, Schematic representation of the CoRE detection algorithm. The genome is partitioned in segments having the same Calder rank (blue representing Inactive, red representing Active regions). The differential rank values (Δr) between the tumour and the control samples were

segmented using a recursive strategy. A null hypothesis of compartment repositioning from replicate comparisons (Δr^{rep}) was derived and tested to compute the significance of each computed segment. The output of the algorithm is a list of activating and in-activating regions. **d**, Number of Mbp covered by activating and inactivating CoRE regions in the three 20 weeks post-WGD tumours. **e**, Distribution of CoRE region sizes (in bp) between activating and inactivating regions in the three 20 weeks post-WGD tumours. **f**, The relationship between differential compartment rank (x-axis) and differential ChIP-seq signal (y-axis) between the tumour and control sample for each tumour and each histone mark. Points represent 50-kb genomic bins. **g**, Expected distributions of correlations (grey) between changes in sub-compartment rank and changes in histone modification signal across different histone marks and different 20 weeks post-WGD tumours in RPE$^{TP53-/-}$ cells. Observed correlations are highlighted as red lines. Empirical *p*-values are shown. *p*-values calculated by data permutation ($n = 1000$) **h**, Percentage of the genome covered by a CoRE in the three RPE$^{TP53-/-}$ 20 weeks post-WGD derived tumour samples. **i**, Example of CoREs detected using the DiffComp or CBS algorithm for delta-rank segmentation. **j**, Distribution of distances between DiffComp and CBS breakpoints (red) and between DiffComp breakpoints and randomly generated ones by shuffling the CBS breakpoints across the genome. *p*-value calculated by Kolmogorov-Smirnov test. **k**, Volcano plot representation of CoREs detected by DiffComp (left) and CBS (right). CoREs involving the selected oncogenes and tumour suppressors discussed in the manuscript are labelled and detected by both approaches. *p*-values calculated by DiffComp. **l**, Number of activating and inactivating CoREs detected by both algorithm (shared) or by only one of them.

Source data

Extended Data Fig. 11 Examples of compartment repositioning events associated with gene expression and epigenetic changes.

a, d, Detailed characterization of histone modification and CTCF changes in the indicated genomic region in RPE$^{TP53-/-}$ control and tumour samples. **b, e**, Detailed characterization of the compartment and histone modification

and CTCF changes in the indicated genomic region in RPE$^{TP53-/-}$ control and tumour samples. Top: sub-compartment assignments inferred by Calder. Bottom: histone mark intensities **c, f**, Significant interactions within the indicated regions in RPE$^{TP53-/-}$ control and tumour 1 samples. Histone mark intensities for the corresponding sample are shown at the bottom. **g,h**, Percentage of activating (*red*, consistent n = 91, inconsistent n = 22) and inactivating (*blue*, consistent n = 170, inconsistent n = 30) CoREs that were detected by CBS and that could be traced back to compartment changes occurring already in WGD cells.

Source data

Extended Data Fig. 12 CP-A$^{TP53-/-}$ cells acquiring spontaneous high ploidy show chromosomal instability and significant compartment repositioning.

a, Flow-cytometry based sorting of the spontaneous high-ploidy cells from the CP-A$^{TP53-/-}$ population. **b**, Inter- chromosomal Hi-C maps of spontaneous high-ploidy CP-A$^{TP53-/-}$ cells show contact patterns consistent with chromosomal translocations (arrows). **c**, Relative copy number profile of control and spontaneous high-ploidy CPA$^{TP53-/-}$ cells (red: gains, blue: losses). **d**, Density distribution of the difference (delta) of compartment ranks of matching genomic bins between of spontaneous high-ploidy and control cell (red), induced WGD and control cells (black), and replicates of control cells (gray). **e,f**, Volcano plots showing the delta compartment rank values for candidate CoREs (x-axis) and associated p-values (y-axis) obtained from the comparison of spontaneous high-ploidy and control cells (**e**) and induced WGD and control cells (**f**). Significant CoREs are in red and blue, non-significant are in gray. **e-f** *p*-values calculated by DiffComp. **g**, Copy number profiles for each colony isolated from soft agar for CP-A$^{TP53-/-}$ samples. **h**, Percentage of consistent and inconsistent inactivating (blue) and activating (red) CoREs for each colony.

Source data

Supplementary information

Supplementary Information

This file contains Supplementary Figs. 1–6.

Reporting Summary

Peer Review File

Supplementary Table 1

Differential expression analysis between RPE *TP53*$^{-/-}$ control cells and WGD cells as determined using the DESeq2 algorithm. *P* values were computed using the DESeq2 package.

Supplementary Table 2

scHi-C quality metrics.

Supplementary Table 3

MAF file of all detected mutations that passed quality filters.

Supplementary Table 4

CNVs detected using Control-FREEC in RPE *TP53*$^{-/-}$ control, WGD and post-WGD samples, in vitro and in vivo.

Supplementary Table 5

Differential expression analysis between RPE *TP53*$^{-/-}$ control and RPE *TP53*$^{-/-}$ 20-weeks post-WGD derived tumours as determined using the MAST algorithm. *P* values were computed using the MAST package.

Supplementary Table 6

List of statistically significant CoREs identified in the three RPE *TP53*$^{-/-}$ 20-weeks post-WGD derived tumour samples and in the CP-A *TP53*$^{-/-}$ colonies. *P* values were computed using the DiffComp algorithm, as detailed in the Methods.

Source data

Source Data Fig. 1

Source Data Fig. 2

Source Data Fig. 3

Source Data Fig. 4

Source Data Fig. 5

Source Data Fig. 6

Source Data Extended Data Fig. 1

Source Data Extended Data Fig. 2

Source Data Extended Data Fig. 3

Source Data Extended Data Fig. 4

Source Data Extended Data Fig. 5

Source Data Extended Data Fig. 6

Source Data Extended Data Fig. 7

Source Data Extended Data Fig. 8

Source Data Extended Data Fig. 9

Source Data Extended Data Fig. 10

Source Data Extended Data Fig. 11

Source Data Extended Data Fig. 12

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Lambuta, R.A., Nanni, L., Liu, Y. *et al.* Whole-genome doubling drives oncogenic loss of chromatin segregation. *Nature* **615**, 925–933 (2023). https://doi.org/10.1038/s41586-023-05794-2

- Received: 13 February 2022
- Accepted: 03 February 2023

- Published: 15 March 2023
- Issue Date: 30 March 2023
- DOI: https://doi.org/10.1038/s41586-023-05794-2

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

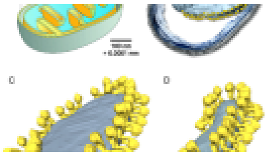
Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

[Genome doubling perturbs DNA packing and promotes cancer development](https://www.nature.com/articles/s41586-023-05794-2)

Nature Research Briefing 15 Mar 2023

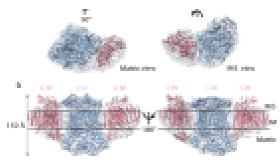
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05794-2



Protonic Capacitor: Elucidating the biological significance of mitochondrial cristae formation

29 June 2020

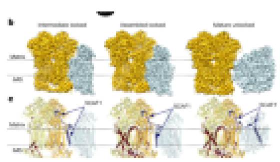
James Weifu Lee



Cryo-EM structure of the yeast respiratory supercomplex

31 December 2018

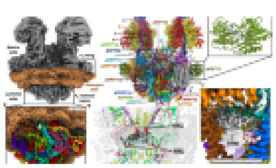
Sorbhi Rathore, Jens Berndtsson, … Martin Ott



Structure and assembly of the mammalian mitochondrial supercomplex CIII2CIV

06 October 2021

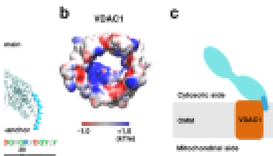
Irene Vercellino & Leonid A. Sazanov



[Type III ATP synthase is a symmetry-deviated dimer that induces membrane curvature through tetramerization]

22 October 2020

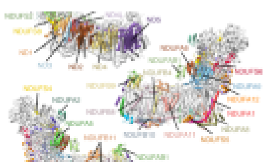
Rasmus Kock Flygaard, Alexander Mühleip, … Alexey Amunts



[Structural basis of complex formation between mitochondrial anion channel VDAC1 and Hexokinase-II]

03 June 2021

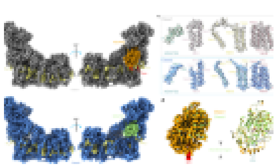
Nandan Haloi, Po-Chao Wen, … Emad Tajkhorshid



[Mitochondrial complex I structure reveals ordered water molecules for catalysis and proton translocation]

03 August 2020

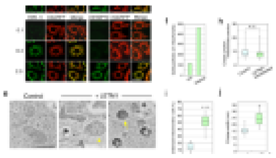
Daniel N. Grba & Judy Hirst



[The coupling mechanism of mammalian mitochondrial complex I]

10 February 2022

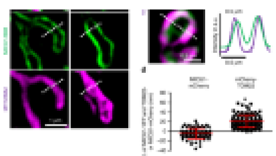
Jinke Gu, Tianya Liu, … Maojun Yang



[The mitochondrial inner membrane protein LETM1 modulates cristae organization through its LETM domain]

05 March 2020

Seiko Nakamura, Aiko Matsui, … Toshihiko Oka



[MICU1 controls cristae junction and spatially anchors mitochondrial Ca2+ uniporter complex]

19 August 2019

Benjamin Gottschalk, Christiane Klec, … Wolfgang F. Graier

- Article
- Open Access
- Published: 22 March 2023

Structural basis of mitochondrial membrane bending by the I–II–III$_2$–IV$_2$ supercomplex

- Alexander Mühleip,
- Rasmus Kock Flygaard,
- Rozbeh Baradaran,
- Outi Haapanen,
- Thomas Gruhl,
- Victor Tobiasson,
- Amandine Maréchal,
- Vivek Sharma &
- …
- Alexey Amunts

Show authors

Nature **volume 615**, pages 934–938 (2023)

- 14k Accesses
- 109 Altmetric
- Metrics details

Abstract

Mitochondrial energy conversion requires an intricate architecture of the inner mitochondrial membrane[1]. Here we show that a supercomplex containing all four respiratory chain components contributes to membrane curvature induction in ciliates. We report cryo-electron microscopy and cryo-tomography structures of the supercomplex that comprises 150 different proteins and 311 bound lipids, forming a stable 5.8-MDa assembly. Owing to subunit acquisition and extension, complex I associates with a complex IV dimer, generating a wedge-shaped gap that serves as a binding site for complex II. Together with a tilted complex III dimer association, it results in a curved membrane region. Using molecular dynamics simulations, we demonstrate that the divergent supercomplex actively contributes to the membrane curvature induction and tubulation of cristae. Our findings highlight how the evolution of protein subunits of respiratory complexes has led to the I–II–III_2–IV_2 supercomplex that contributes to the

shaping of the bioenergetic membrane, thereby enabling its functional specialization.

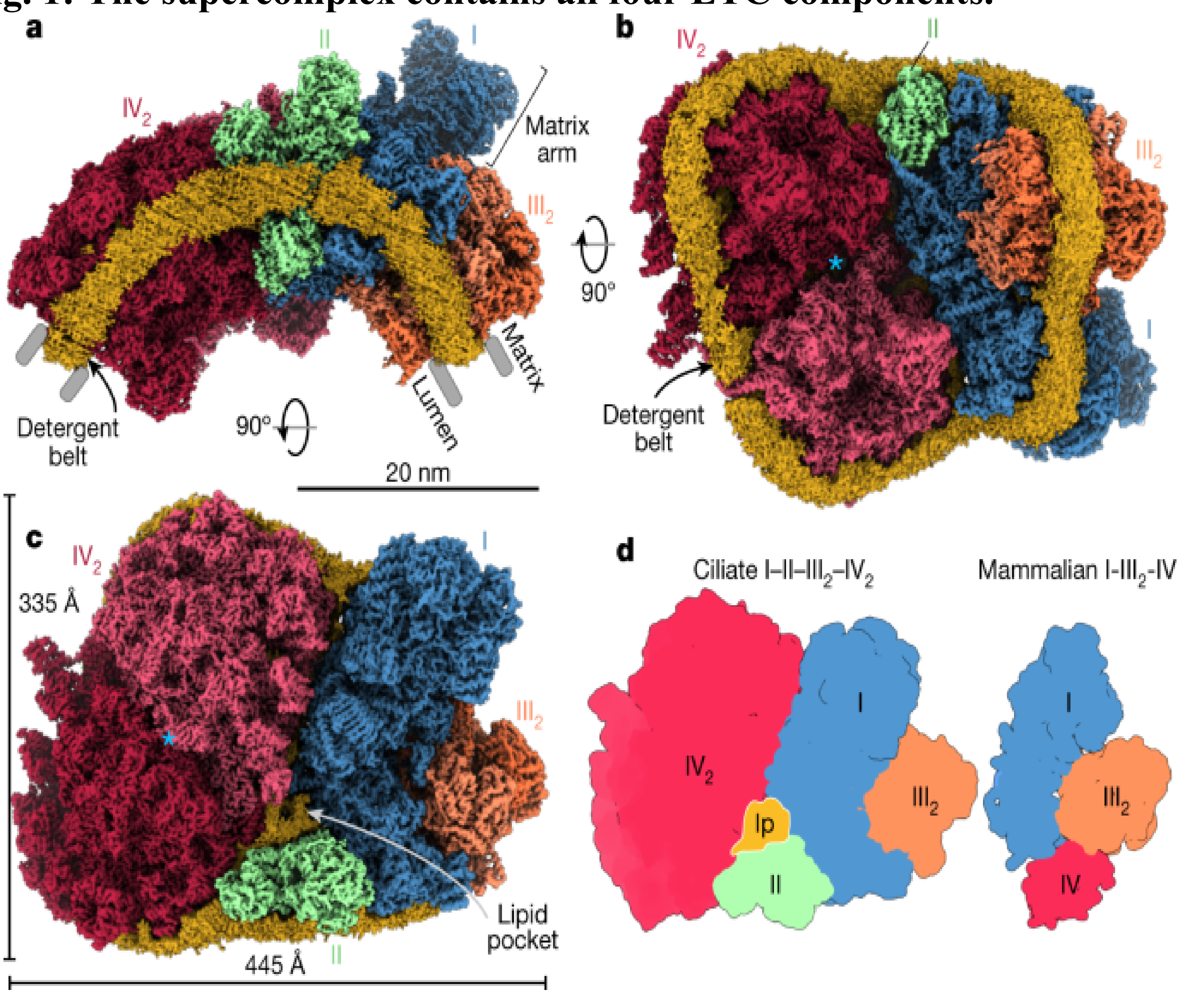
Main

Mitochondrial energy conversion requires an electron transport chain (ETC) that generates a proton motive force across the inner mitochondrial membrane to drive the essential adenosine triphosphate (ATP) formation by F_1F_o-ATP synthase. The ETC consists of four multisubunit membrane complexes: complex I (CI; NADH:ubiquinone (UQ) oxidoreductase), complex II (CII; succinate:UQ oxidoreductase), complex III (CIII; cytochrome bc_1) and complex IV (CIV; cytochrome *c* oxidase). Biochemical and structural analyses have shown that these components can organize into supercomplexes containing CI, CIII dimer ($CIII_2$) and CIV[1,2]. CII transfers electrons from succinate via its covalently bound flavin adenine dinucleotide (FAD) and iron–sulfur clusters to UQ and is also a component of the tricarboxylic acid cycle, making a functional link between the two central metabolic pathways[3]. Although CII has been suggested to interact with mammalian ETC complexes[4,5,6,7,8], it was not experimentally found as part of any characterized supercomplex. In addition, for the bioenergetic process to occur, a specific topology of the crista membranes that form functionally distinct high-potential compartments is critical[9]. An established mechanism for maintenance of such a topology relies on oligomerization of ATP synthase and its specific interplay with lipids[10,11,12,13,14]. In ciliates, the inner mitochondrial membrane is organized as tubular cristae, which was previously explained by the helical row assembly of ATP synthase[12,15,16].

We purified the intact respiratory supercomplex from mitochondria of the ciliate protist *Tetrahymena thermophila* and determined its structure by single-particle cryo-electron microscopy (cryo-EM) (Extended Data Fig. 1 and Supplementary Table 1). At an overall resolution of 2.9 Å, the structure revealed that CI, CII, $CIII_2$ and a CIV dimer (CIV_2) associated into a 5.8-MDa supercomplex (Fig. 1). When viewed along the membrane plane, the assembly of more than 300 transmembrane helices displays a bent shape, indicating that the accommodating membrane adopts a local curvature with a radius of approximately 20 nm (Fig. 1). Masked refinements resolved

individual structures that together form an assembly of 150 different protein subunits and 311 bound lipids (Extended Data Figs. 1 and 2a–f and Supplementary Tables 1 and 2). CIV_2 is associated with the long side of the membrane region of CI, opposite $CIII_2$. This arrangement is markedly different compared with known mammalian supercomplexes[17,18] and correlates with the acquisition of four ciliate-specific CI subunits that would clash with the position of CIV as seen in mammals (Fig. 1d and Extended Data Fig. 3). CII is anchored between CI and CIV, highlighting the unique architecture and composition of the native supercomplex (Fig. 1). In-gel activity assays confirmed co-migration of functional CI, CII and CIV in a high-molecular-mass band, which we assigned to the intact supercomplex (Extended Data Fig. 2g; and see source data for Extended Data Fig. 2).

Fig. 1: The supercomplex contains all four ETC components.

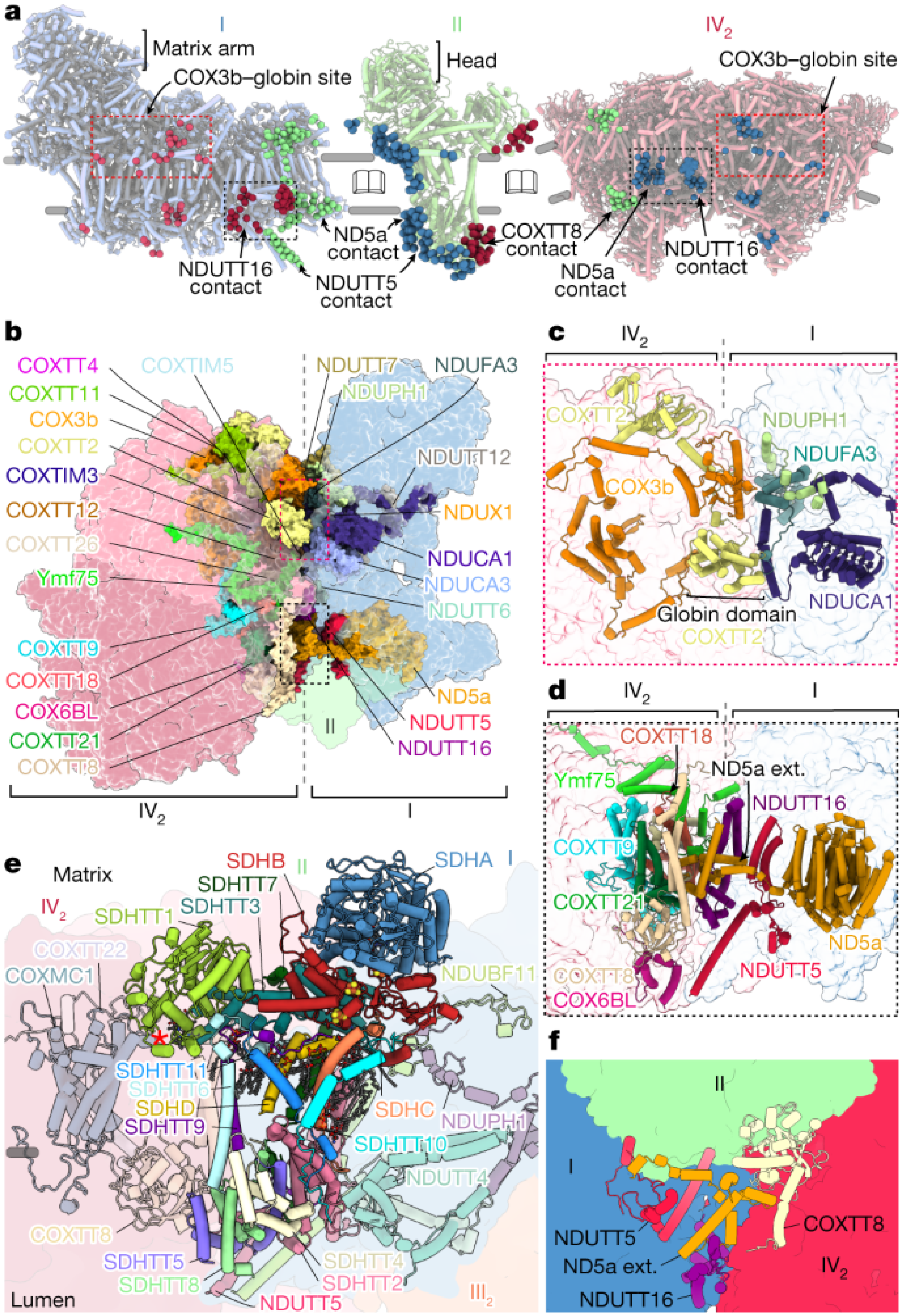


a, Side view of the supercomplex density showing the curved detergent micelle (yellow). **b**, Lumenal view illustrates how complexes CI, CII and CIV_2 stabilize each other. The blue asterisk indicates the symmetry axis of CIV_2. **c**, Matrix view shows CII binding in a wedge between CI and CIV_2, resulting in the enclosure of a lipid pocket (lp). The blue asterisk indicates the same position as in **b**. **d**, Architecture comparison of the ciliate I–II–III_2–IV_2 supercomplex (this study; left) with the mammalian I–III_2–IV respirasome (Protein Data Bank (PDB) accession 5J4Z; right), highlighting a different location of CIV_2 that is correlated with acquisition of CI subunits that stabilize $CIII_2$.

CIV_2 is the most divergent of the four ETC complexes (Extended Data Figs. 4 and 5). We modelled 105 lipids, four UQs and 53 protein chains per monomer, of which four are mitochondrially encoded. We found that two of those subunits, previously annotated as ciliate-specific Ymf67 and Ymf68 (also known as COX3) represent complementary protein fragments, with coding genes split in the mitochondrial genome by a $tRNA^{Trp}$ gene insertion (Extended Data Fig. 6). Each fragment is extended by over 400 and 200 residues, respectively. Together, they form a functional COX3 (COX3a and COX3b), including the conserved seven transmembrane helix fold (Extended Data Fig. 6e). In our structure, COX3a and COX3b extend throughout the CIV membrane region, and COX3b has evolved interactions with CI subunits on the matrix side, thereby mediating the supercomplex assembly (Fig. 2a,b). In particular, COX3b forms a contact with a peripheral amphipathic helix of NDUCA1, which is part of a zinc-free γ-carbonic anhydrase (γ-CA) heterotrimer (Fig. 2c). The γ-CA heterotrimer was previously reported in viridiplantae and ciliates (both diaphoretickes)[8,19] and our structure demonstrates that it acts as a structural scaffold within the supercomplex architecture. Another CI–CIV contact at the same site involves COXTT2 with a N-terminal globin-like domain that interacts with NDUFA3 and was not resolved in the individual CIV_2 structure[8], suggesting that it becomes ordered to mediate supercomplex formation (Fig. 2a,c). The second major interaction site at the CI–CIV interface, which is on the lumenal side of the membrane, also involves a fragmented protein subunit, this time from CI. Consistent with the observation with respect to the protein splitting in CIV, we modelled the N-terminal extension of the ND5 fragment

(ND5a), as well as the newly identified protein subunit NDUTT16 (from CI) (Fig. 2a,b,d). NDUTT16 engages in interactions with at least four subunits of CIV, as well as an interfacial CIV haem group (Fig. 2d and Extended Data Fig. 5f).

Fig. 2: The CI–CIV association and binding of CII.


a
I
II
IV2
Matrix arm
COX3b–globin site
Head
COX3b–globin site
ND5a contact
NDUTT16 contact
NDUTT5 contact
COXTT8 contact
ND5a contact
NDUTT16 contact
b
COXTT4
COXTT11
COX3b
COXTT2
COXTIM3
COXTT12
COXTT26
Ymf75
COXTT9
COXTT18
COX6BL
COXTT21
COXTT8
COXTIM5
NDUTT7
NDUPH1
NDUFA3
NDUTT12
NDUX1
NDUCA1
NDUCA3
NDUTT6
ND5a
NDUTT5
NDUTT16
II
IV2
I
c
IV2
I
COXTT2
NDUPH1
NDUFA3
COX3b
NDUCA1
Globin domain
COXTT2
d
IV2
I
COXTT18
Ymf75
ND5a ext.
NDUTT16
COXTT9
COXTT21
ND5a
COXTT8
COX6BL
NDUTT5
e
Matrix
IV2
SDHB
II
SDHA
I
SDHTT7
SDHTT3
SDHTT1
COXTT22
COXMC1
NDUBF11
SDHTT11
SDHTT6
SDHD
SDHTT9
SDHC
NDUPH1
SDHTT10
NDUTT4
COXTT8
SDHTT5
SDHTT8
SDHTT4
SDHTT2
NDUTT5
III2
Lumen
f
II
I
NDUTT5
ND5a ext.
NDUTT16
COXTT8
IV2

a, Open-book view with contact sites of CI with CII–CIV_2 (left), CII with CI–CIV_2 (middle) and CIV_2 with CI–CII (right). Interactions are shown as spheres (CI in blue, CII in green, CIV_2 in red or dark pink). Only one CIV monomer interacts with CII. The main interaction sites are indicated. **b**, The CI–CIV_2 interacting subunits are shown in coloured surfaces. **c**, The COX3b–NDUCA1 contact site. **d**, The NDUTT16–ND5a contact site. ext., extension. **e**, CII binding to subunits of CI–CIV (transparent). The red asterisk marks C-type haem. **f**, CI, CII and CIV are connected via the membrane and lumenal regions.

Our finding of the split core subunits gaining a capacity of establishing intercomplex contacts to stabilize the supercomplex that curves the membrane suggests a putative mechanism by which gene fragmentation, followed by its expansion, might convey subunit function, which was also observed in mitoribosomes[20,21]. Overall, the CI–CIV_2 interface involves 25 subunits, forming an extensive buried interface of approximately 2,300 $Å^2$ with a curved membrane region (Fig. 2a,b), and thus a single mutation is unlikely to disrupt the curvature induction.

T. thermophila CII binds in a wedge-shaped gap formed by CI–CIV in our structure (Fig. 1a–c). In addition to the four canonical subunits (SDHA–D), it is composed of 11 ciliate-specific subunits SDHTT1–11 (Fig. 2e and Extended Data Fig. 7). The matrix module SDHA and SDHB forms a conserved head region, containing both the covalently bound FAD and the three iron–sulfur clusters (Extended Data Fig. 7b). The membrane anchor is formed by two small subunits, SDHC (7 kDa) and mitochondria-encoded SDHD (5 kDa), which could only be assigned by locating topologically conserved transmembrane helices in the map (Extended Data Fig. 7a,c,d). At the lumen, an approximately 70-kDa module (SDHTT2, SDHTT4, SDHTT5 and SDHTT8) anchors CII to CI–CIV (Fig. 2e,f and Extended Data Fig. 7a). SDHTT5 interacts with a helix of NDUTT5 protruding from the membrane arm of CI, and with the Surf1-like protein subunit COXTT8 (CIV) (Fig. 2a,e). At the same position, COXTT8 interacts with CI via the N-terminal ND5a extension, and with Ymf75, COXTT27 and COXTT18 at the CIV dimer interface. Thus, the three complexes—CI, CII and CIV—are connected in the lumen (Fig. 2f).

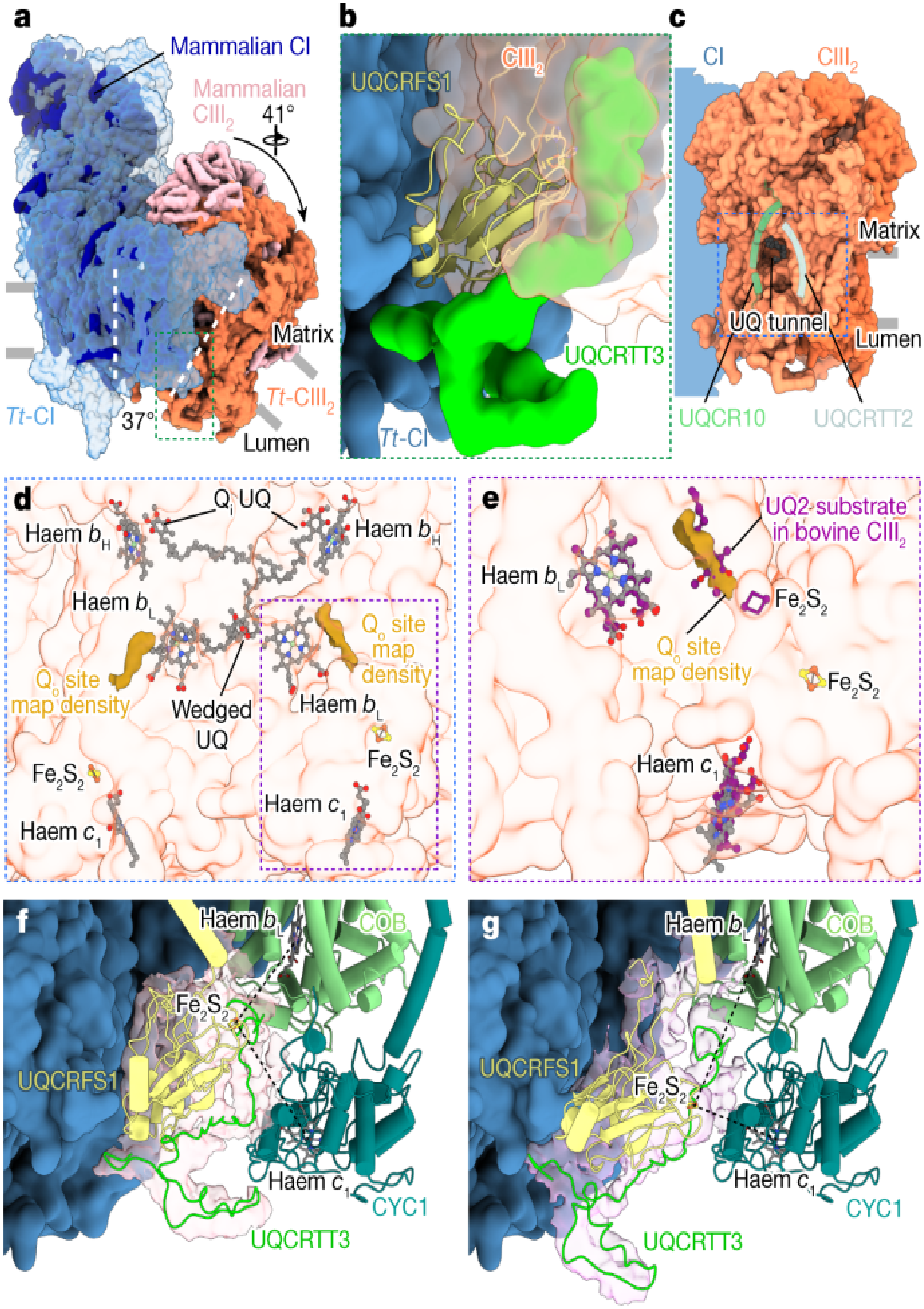
Between SDHB and SDHD, we identified a ligand, which we assign to ubiquinone bound to the conserved proximal Q_p site[22,23] (Extended Data Fig. 7e). On the matrix side, the 36-kDa soluble subunit SDHTT1 contains a bis-histidine C-type haem covalently bound by a single cysteine residue (Extended Data Fig. 7a). Although it is exposed to the membrane region, at a distance of approximately 60 Å from the Fe_3S_4 cluster, the non-canonical haem C is located too far to participate in direct CII electron transfer (Extended Data Fig. 7a). To provide further evidence of the presence of additional haem groups in the supercomplex, we recorded visible absorption redox spectra of the purified sample (Extended Data Fig. 8. Deconvolution of the merged absorption bands characteristic of B-type and C-type haems clearly revealed the contribution of at least one additional haem group in the supercomplex, in addition to those present in CIII, with specific absorption at 556 nm.

The presence of a functional ETC with CII is consistent with previous observations that *T. thermophila* can utilize succinate to drive cellular respiration[24]. Our native structure with bound CII, which contributes to the ubiquinol pool, demonstrates that assembly of the supercomplex is not limited to the proton-pumping respiratory complexes (CI, CIII and CIV). Beyond decreasing cytochrome *c* transfer distance[25], this suggests a potential role of supercomplex formation in mediating increased UQ diffusion, as suggested in analogous membrane systems with high protein to lipid ratios[26,27]. Furthermore, the tubular membrane morphology may require anchoring of CII to the curved supercomplex to retain it in the functionally relevant cristae and prevent diffusion to flat membrane regions.

$CIII_2$ in our structure is tilted with respect to CI by 37° (Fig. 3a). This tilted arrangement offsets the transmembrane region, consistent with its curved membrane environment. The interface involves 20 subunits and 19 bound native lipids interacting within the matrix, transmembrane and lumenal domains (Extended Data Fig. 9). When compared with the mammalian counterpart, $CIII_2$ is rotated by 41° and shifted approximately 14 Å due to acquisition of four CI subunits, as well as the CIII subunit UQCRTT1 (refs. [17,28,29,30]) (Fig. 3a and Extended Data Fig. 9a–c). This arrangement results in a specific CI–$CIII_2$ contact with one copy of the Rieske iron–sulfur protein

of CIII (UQCRFS1) interacting with the CI membrane arm (Fig. 3b). The interaction site is further augmented by a hitherto unidentified protein UQCRTT3, which interacts with the lumenal head domain of UQCRFS1 and wedges in between the interface to CYC1 (Fig. 3b).

Fig. 3: An altered CI–$CIII_2$ interface maintains functional symmetry of $CIII_2$.


a
Mammalian CI
Mammalian CIII$_2$
41°
37°
Matrix
Lumen
Tt-CI
Tt-CIII$_2$
b
UQCRFS1
CIII$_2$
UQCRTT3
Tt-CI
c
CI
CIII$_2$
Matrix
Lumen
UQ tunnel
UQCR10
UQCRTT2
d
Q$_i$ UQ
Haem b_H
Haem b_H
Haem b_L
Q$_o$ site map density
Q$_o$ site map density
Wedged UQ
Haem b_L
Fe$_2$S$_2$
Fe$_2$S$_2$
Haem c_1
Haem c_1
e
UQ2 substrate in bovine CIII$_2$
Haem b_L
Fe$_2$S$_2$
Q$_o$ site map density
Fe$_2$S$_2$
Haem c_1
f
Haem b_L
COB
Fe$_2$S$_2$
UQCRFS1
Haem c_1
CYC1
UQCRTT3
g
Haem b_L
COB
Fe$_2$S$_2$
UQCRFS1
Haem c_1
CYC1
UQCRTT3

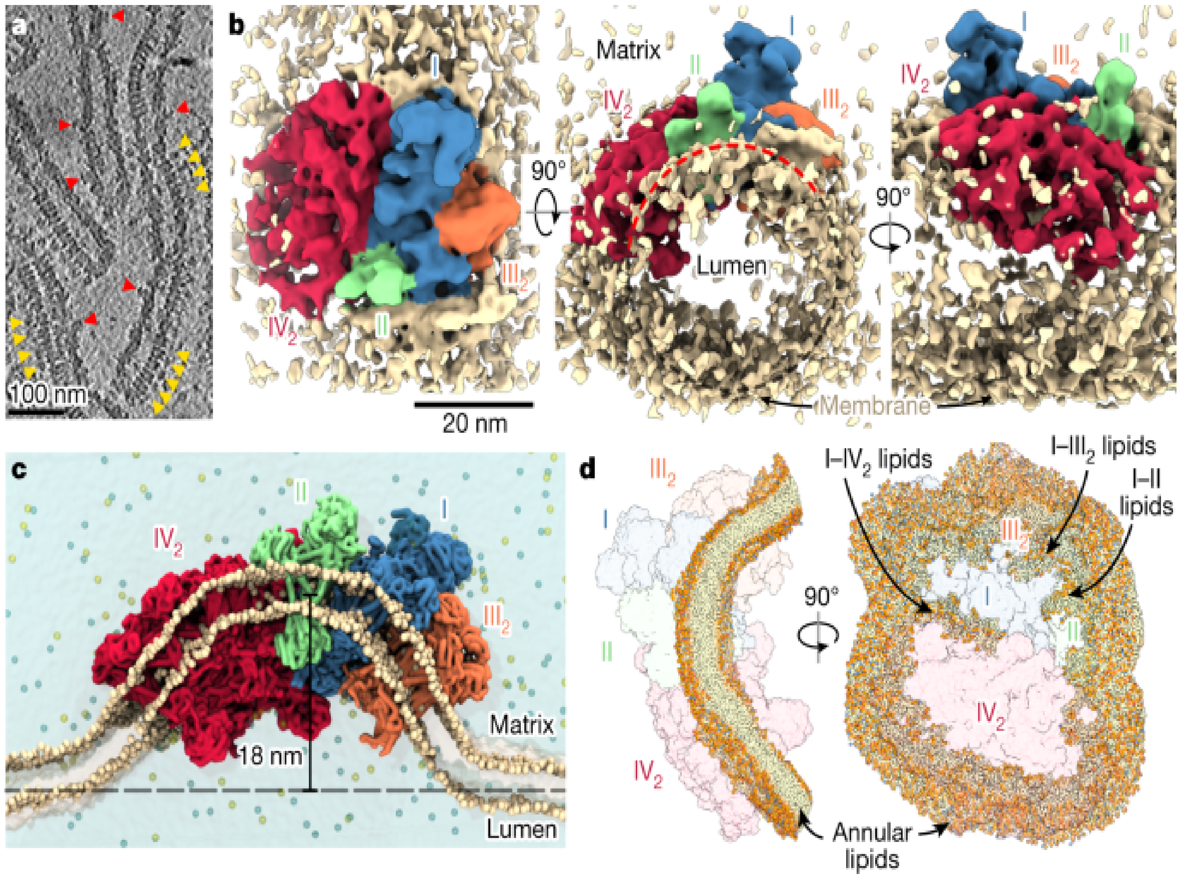
a, Superposition of *T. thermophila* CI–$CIII_2$ (*Tt*-CI–*Tt*-$CIII_2$) with mammalian CI–$CIII_2$ (PDB accession 5J4Z) shows that *T. thermophila* $CIII_2$ is tilted, rotated and displaced with respect to the CI membrane arm, as indicated by the white dashed lines and arrow, due to acquisition of new proteins. The green box view is shown in **b**. **b**, Tilted *T. thermophila* $CIII_2$ results in an interaction between UQCRFS1 and the CI membrane arm together with UQCRTT3. **c**, *T. thermophila* $CIII_2$ shows a membrane-accessible tunnel extending to the COB Q_i sites. The blue box view is shown in **d**. **d**, The Q_i UQs are located close to haem b_H, with one UQ wedged in between the two haem b_L molecules. The density map of *T. thermophila* $CIII_2$ shows two features corresponding to Q_o sites. The purple box view is shown in **e**. **e**, The Q_o site density overlaps with ubiquinol in the bovine $CIII_2$ structure (PDB accession 1NTZ). **f**,**g**, 3D maps showing B state (**f**) and C state (**g**) conformations of UQCRFS1 and UQCRTT3. Only COB and CYC1 proteins are shown for clarity, highlighting distances of Fe_2S_2 to haem c_1 and b_L (black dashes).

We traced the membrane-accessible UQ tunnel, lined by UQCR10 and ciliate-specific subunit UQCRTT2 (Fig. 3c), leading to the conserved COB haem b_H, where the density for a bound (semi)-UQ was observed in the Q_i site[31] (Fig. 3d and Extended Data Fig. 9e,f). Furthermore, we observed map density features close to the two conserved haem b_L groups, which probably correspond to ubiquinols bound at the Q_o sites (Fig. 3d,e and Extended Data Fig. 9e,f). The distances between the Q_o site, haem b_L and haem b_H within each CIII monomer are consistent with those observed in mammalian $CIII_2$, with the two haem b_L in COB being bridged by a non-canonical UQ. We detected density for two copies of the flexible UQCRFS1 head domain (Extended Data Fig. 9g), which contrasts with recent work that found only the head domain proximal to the CI quinone tunnel to display flexibility, whereas the distal domain at the CI interface was proposed to be non-functional in electron transport[8]. Using focused 3D classification for the distal UQCRFS1 head domain, we then identified two classes probably representing the extremes of the head domain movement: from the B state where the Fe_2S_2 cluster is distanced from haem c_1 to the C state where the

Fe_2S_2 cluster is closest[32,33] (Fig. 3f,g and Extended Data Fig. 9i). This movement of the UQCRFS1 head domain is coupled to conformational changes in the unidentified UQCRTT3 protein, suggesting a potential role for this subunit in regulation of $CIII_2$ activity. Thus, our observation of the distal UQCRFS1 head domain flexibility, together with the haem b_L-wedged UQ, suggests that functional symmetry is maintained in the ciliate $CIII_2$ despite deviation from the structural symmetry.

To investigate whether the membrane-bending capacity of the supercomplex is biologically relevant, we performed cryo-electron tomography of isolated mitochondrial membranes. Cryo-tomograms revealed approximately 40-nm tubular cristae densely packed with helical ATP synthase rows and supercomplexes, identified by the conspicuous CI matrix arm (Fig. 4a). To elucidate the supercomplex architecture in situ, we performed subtomogram averaging and obtained a map at 28 Å resolution (Extended Data Fig. 1e). The subtomogram average confirmed the presence of the supercomplex, which fits our atomic model (Extended Data Fig. 1f), suggesting that this is the most abundant form. The appearance of a tubular membrane density in the subtomogram average suggests that the supercomplex adopts a preferred orientation, with its CI–CIV_2 interface approximately aligned with the long axis of the tube (Fig. 4b). Furthermore, the curved membrane region of the supercomplex subtends an angle of approximately 130°, indicating that it contributes to the tubular shape of the cristae.

Fig. 4: In situ structure and of the I–II–III_2–IV_2 supercomplex indicate a membrane-bending function.



a, Cryo-tomographic slice of tubular cristae with a diameter of approximately 40 nm. ATP synthase and the supercomplex are marked with yellow and red arrowheads, respectively. **b**, Subtomogram average of the I–II–III_2–IV_2 supercomplex revealing a preferred orientation in tubular membranes and an arc-shaped structure subtending approximately 130° (red dashes). **c**, Coarse-grained molecular dynamics simulation showing that the arched membrane region of the supercomplex generates a significant membrane curvature, resulting in an 18-nm local displacement of the membrane from the bilayer plane. **d**, Molecular dynamics simulation reveals the curved structure of the annular lipid shell surrounding the supercomplex, as well as lipid-filled subcomplex interfaces.

To elucidate the membrane-shaping activity of the supercomplex, we performed coarse-grained molecular dynamics simulations. When placed into a planar lipid bilayer, the supercomplex induces a curved membrane topology, displacing the membrane by 18 nm from the original plane (Fig. 4c and Supplementary Video 1), in contrast to the protein-free lipid bilayer,

which remains relatively flat (Supplementary Video 2). Furthermore, the annular lipid shell surrounding the complex in the equilibrated system displays a highly curved architecture, supportive of an active role in membrane curvature induction (Fig. 4d and Supplementary Video 3). In addition, we observed lipid pockets in the transmembrane interfaces between subcomplexes, which suggests that their maintenance is crucial for the integrity of the supercomplex (Extended Data Fig. 2a,b), as was also shown for photosynthetic membranes[34]. Molecular dynamics simulations of CIV_2 in a membrane suggest that it can also induce membrane bending in its immediate vicinity, similar to the supercomplex (Extended Data Fig. 10a,b). However, for the formation and stability of a tubular architecture, juxtapositioning of individual subcomplexes is probably necessary. This is also supported by the increased arc length of the detergent belt observed in the cryo-EM structure of the supercomplex, when compared with CIV_2 (Extended Data Fig. 10b). Finally, molecular dynamics simulations of the supercomplex lacking CII showed a similar wrapping of the lipid bilayer as for the full supercomplex, with lipids filling the generated void (Extended Data Fig. 10c). This indicates that CII association may contribute to complex stability while retaining the enzyme in the tubular cristae regions, where ubiquinol is required as a substrate.

Our results indicate that cristae shaping involves both the respiratory supercomplex and the ATP synthase that together generate membrane tubulation. Although the coiled ATP synthase rows fix the helix diameter at 130 nm (refs. [12,15]), supercomplexes serve the function of confining a narrow crista diameter of around 40 nm, which allows tight packing of cristae, thereby increasing the surface area of the bioenergetic membrane. This membrane-shaping organization of the respiratory supercomplex is markedly different from the mammalian homologue, which resides in the flat crista regions. Furthermore, because every crista represents an independent functional compartment[35], restriction of the crista diameter by the respiratory supercomplex probably serves to minimize the volume, thereby potentially contributing to higher local concentration of electron carriers, ensuring that proton translocation results in an increased local proton motive force, ultimately optimizing conditions for ATP synthesis. This is consistent with the observation that mutant yeast strains with large, balloon-like cristae display respiratory defects[36,37]. Thus, our findings show

how respiratory supercomplexes together with other factors can organize the architecture of the bioenergetic membrane, providing a mechanism for enabling its functional specialization.

Methods

Purification of *T. thermophila* supercomplex

T. thermophila cells were grown at 36 °C and harvested as previously described[12,20], and cell pellets were resuspended in homogenization buffer (20 mM HEPES-KOH pH 7.5, 350 mM d-mannitol, 5 mM EDTA and 1x protease-inhibitor tablet) and lysed in a Dounce homogenizer on ice. Intact mitochondria were isolated first by differential centrifugation of the lysate; following membrane solubilization, the lysate was cleared at 30,000*g* for 30 min at 4 °C, and finally, on a discontinuous sucrose gradient with 15%, 23%, 32% and 60% w/v sucrose in buffer SEM (20 mM HEPES-KOH pH 7.5, 250 mM sucrose and 1 mM EDTA) at 14,1371*g* for 60 min at 4 °C in an SW28 rotor. Intact mitochondria sedimented to the interface between 32% and 60% sucrose and were collected from the gradient, snap-frozen in liquid nitrogen and stored at −80 °C. The isolated mitochondria were lysed in buffer A (25 mM HEPES-KOH pH 7.5, 25 mM KCl, 5 mM $MgCl_2$ and 4% w/v digitonin) for 1 h on ice. This procedure was previously confirmed as a gentle solubilization method. Following mitochondrial membrane solubilization, cleared lysate was placed on a sucrose cushion (25 mM HEPES-KOH pH 7.5, 25 mM KCl, 5 mM $MgCl_2$, 0.1% w/v digitonin and 30% w/v sucrose) in Ti70 tubes and centrifuged at 164,685*g* for 1 h. The pellet was gently washed and finally resuspended in buffer D (25 mM HEPES-KOH pH 7.5, 25 mM KCl, 5 mM $MgCl_2$ and 0.1% w/v digitonin). Before loading sample material on a size-exclusion chromatography column, larger aggregates were pelleted at 30,000*g* for 20 min at 4 °C. Cleared sample was loaded on a Superose 6 Increase 3.2/300 column equilibrated in buffer D, collecting elution fractions of 100 µl throughout the run. Peak fractions were used immediately for cryo-grid preparation.

UV–visible difference spectroscopy

The sample obtained from the sucrose cushion step was analysed for haem content and supercomplex composition using UV–visible difference spectroscopy and clear native-PAGE (CN-PAGE) combined with in-gel activity assays. UV–visible difference spectra were recorded between 390 nm and 675 nm using a home-built spectrophotometer. Protein samples were diluted as necessary in 50 mM HEPES and 0.1% digitonin at pH 8.0. Spectra were measured from sodium dithionite-reduced minus air-oxidized spectra. When multiple absorption bands overlapped, spectra were deconvoluted using the peak analysis function in OriginPro 2015 (OriginLab Corporation).

Gel electrophoresis

NativePAGE 3 to 12%, Bis-Tris, 1.0 mm, Mini Protein Gel (Invitrogen) pre-cast gels were used for CN-PAGE. The gels were loaded with a protein ladder (NativeMark, Invitrogen) and four identical sample lanes where the protein samples were mixed with NativePAGE Sample Buffer (final concentration, 50 mM BisTris, 6N HCl, 50 mM NaCl, 10% w/v glycerol and 0.001% Ponceau S, pH 7.2) as per the manufacturer’s instruction. Electrophoresis was conducted at 4 °C, first at 150 V for 30 min with NativePage Light Blue Cathode buffer (50 mM BisTris, 50 mM Tricine, pH 6.8, and 0.002% Coomassie G-250) and then at 250 V for 150 min with NativePAGE Anode Buffer (50 mM BisTris and 50 mM Tricine, pH 6.8). In-gel activity assays were performed following published protocols[38]. In brief, each sample lane from CN-PAGE was incubated with an aqueous solution to reveal (1) protein bands (0.02% Coomassie G-250, overnight) or the presence of active (2) CI (2 mM Tris pH 7.4, 2.5 mg ml^{-1} nitrotetrazolium blue chloride (NBT) and 0.1 mg ml^{-1} NADH, 15 min), (3) CII (5 mM Tris pH 7.4, 2.5 mg ml^{-1} NBT, 84 mM succinic acid and 0.2 mM phenazine methosulfate, 30–40 min) and (4) CIV (0.05 mM KPi pH 7.4, 0.5 mg ml^{-1} 3,3′-diaminobenzidine (DAB) and 1 mg ml^{-1} cytochrome *c* from *Saccharomyces cerevisiae*, overnight). The reactions were stopped by incubation in 10% (v/v) acetic acid, followed by multiple exchanges of water. The gels shown in this paper are representative of two experiments from two separate supercomplex preparations.

Cryo-EM sample preparation and data collection

Supercomplex eluted at a concentration of approximately 10 mg ml^{-1}. Aliquots of the peak fraction were diluted in buffer D to 0.75 mg ml^{-1} before applied to cryo-grids. Quantifoil R2/2-300 grids floated with a homemade 3-nm amorphous carbon layer were glow-discharged immediately before applying a 3 µl sample. Grids were vitrified using liquid ethane cooled by liquid nitrogen in a Vitrobot Mark IV, with a 30-s wait time before blotting grids for 3 s at blot force of 0. A total of 26,063 movies were collected on a Titan Krios (Thermo Fisher Scientific) operated at 300 kV at a nominal magnification of ×165,000 (0.83 Å per pixel) with a Quantum K2 camera (Gatan) using a slit width of 20 eV. An objective lens aperture of 70 µm was used, with an exposure rate of 4.26 electrons per pixel per second with 5-s exposure fractionated into 20 frames, and the defocus range was 0.6–2.6 µm.

Cryo-EM data processing

Motion correction was performed in the internal implementation of RELION-3.1 (ref. [39]), followed by contrast transfer function estimation by CTFFIND4. Initial rounds of particle picking and 2D classification, followed by ab initio reconstruction, 3D classification and preliminary refinement of the supercomplex. Template-based particle picking in RELION was then used to pick and extract 1,664,103 particles. 2D classification and 3D heterogeneous refinement steps in cryoSPARC v2 (ref. [40]) were then used to separate supercomplex particles from copurified ATP synthase, resulting in a final 138,746 supercomplex particles used for subsequent refinement. Following a consensus refinement in cryoSPARC, per-particle contrast transfer function refinement and Bayesian polishing were performed in RELION-3.1. For final refinements in cryoSPARC, particles were downsampled from a 724-pixel box to 480 pixels, resulting in a pixel size of 1.25 Å per pixel. Masked refinements of the respective supercomplex subregions resulted in map resolutions of 2.9 Å for the entire supercomplex, 2.8 Å for CI, 3.0 Å for CII, 2.8 Å for CIII and 2.6 Å for CIV_2. Reported map resolutions are according to gold-standard Fourier shell correlation using the 0.143-criterion. To assess flexibility of the Rieske subunit wedged in the CI–$CIII_2$ interface, we performed focused 3D classification in RELION-3.1 using pre-aligned particles with a mask on the extended area around the headgroup of the Rieske subunit. Classification into ten classes resulted in maps confirming flexibility of the structural

element, with two classes corresponding closely to the previously reported B state and C state (Fig. 3f,g).

Cryo-electron tomography and subtomogram averaging

Crude mitochondrial pellets were resuspended in an equal volume of buffer containing 20 mM HEPES-KOH pH 7.4, 2 mM EDTA, 250 mM sucrose and mixed in a 1:1 ratio with 5-nm colloidal gold solution (Sigma Aldrich) and vitrified as described above on glow-discharged Quantifoil R2/2 Au 200 mesh grids. Tilt series were acquired on a Titan Krios operated at 300 kV with a K3 camera (slit width of 20 eV) using SerialEM or the EPU software (Thermo Fisher Scientific). Mitochondrial membranes were imaged at a nominal magnification of ×42,000 (2.11 Å per pixel) and an exposure rate of 19.5 electrons per pixel per second with a 3 e^{-} $Å^{-2}$ exposure per tilt fractionated into five frames, with tilt series acquired using the exposure-symmetric scheme[41] to ±60° tilt and a 3° tilt increment. Following motion correction in motionCor2, tomographic reconstruction from the tilt series was performed in IMOD[42] using phase-flipping and a binning factor of 2. Tomograms were contrast enhanced using nonlinear anisotropic diffusion filtering to facilitate manual particle picking of supercomplex particles based on the matrix arm of CI. Subtomogram averaging was performed in PEET[43]. Initial references were generated from the data by averaging after rotating subvolumes into a common orientation with respect to the membrane based on manually assigned vectors. Following initial rounds of averaging to generate a suitable reference, data were manually split into half-sets and refined independently, following low-pass filtering to 50 Å. Averaging of 360 particles from 12 tomograms resulted in a 28 Å subtomogram average.

Model building and refinement

Manual model building was performed in Coot[44] and new subunits identified directly for the cryo-EM map. For identified canonical subunits, homology models were generated using SWISS-MODEL. Bound cardiolipins were unambiguously identified from their head group density. Other natively bound lipids were tentatively modelled as phosphatidylcholine, phosphatidylethanolamine or phosphatidic acid based on head group densities. Real-space refinement of atomic models was

performed in PHENIX using secondary structure restraints[45]. Atomic model statistics were calculated using MolProbity[46].

Given the mild solubilization conditions we used, for $CIII_2$, the cryo-EM map showed density located on the pseudo-C_2 symmetry axis between the two COB haem b_L molecules, displaying planar map features consistent with the quinone moiety of UQ. The density clearly indicates that UQ can bind in two orientations, related by the symmetry rotation of the dimer. In either orientation, the quinone moiety is positioned close to a haem b_L, where it potentially could accept electrons for transfer across the dimer axis. In the recent amphipol $CIII_2$ structure[8], the isoprenoid tail of UQ was modelled in the equivalent position; however, the planar density for the quinone was missing. This orientation-equivalent binding of UQ between the two COB haem b_L molecules, together with the B-state and C-state Rieske conformations, suggest a maintained functional symmetry of ciliate $CIII_2$ within the supercomplex.

In CI, we identified 49 canonical subunits and 21 subunits that we assign as phylum-specific. In each CIV monomer, we identified 11 subunits homologous to mammalian CIV (COX1, COX2, COX3a, COX3b, COX5B, COX6A, COX6B, COX6C, COX7A, COX7C and NDUFA4) and 42 ciliate-specific subunits, most of which are peripherally associated around the mitochondrial protein core. Three of the mammalian subunits missing in *T. thermophila* CIV (COX4, COX7B and COX8) are at the interface where two mitochondrial carriers are bound. The mitochondrially encoded core subunit COX3 is split into two fragments. Most of the TM helices are contributed by the C-terminal of COX3b, which is encoded by the mitochondrial *ymf68* gene. The newly annotated Ymf68 is structurally conserved, apart from the missing helix (H1), which is structurally replaced by Ymf67. We therefore assign *ymf67* and *ymf68* of the ciliate mitochondrial genome as separately encoding the COX3a and COX3b subunit fragments, respectively. On the *T. thermophila* mitochondrial genome, *ymf67* and *ymf68* genes are located on the same strand, but are separated by the gene encoding $tRNA^{Trp}$, suggesting that a transposition event may have led to the fragmentation of the original gene encoding COX3. tRNA genes are known to be among the most motile elements in metazoan mitochondrial DNA. Both COX3a and COX3b

fragments have evolved substantial subunit extensions that thread through the augmented CIV monomer unit to recruit lineage-specific subunits and mediate supercomplex assembly.

In the CIV dimer, the dimer interface of 17,000 $Å^2$ is dominated by 16 species-specific subunits. Furthermore, when aligned on the CIV core, a comparison of the mammalian and ciliate structures reveals that the two dimers display markedly different architectures, dimer axes and distances between COX1 cores. This suggests that the dimerization of the ciliate CIV probably evolved through the acquisition of lineage-specific subunits and reflects the constraints of the unique tubular membrane environment.

In addition, each CIV monomer complex contains two different Surf1-like proteins, which were reported to complement defects causing Leigh syndrome in humans. In our structure, two Surf1-like proteins are permanently attached to CIV and display similar overall structures, consisting of a lumen-exposed soluble domain and a TM helix hairpin. The two Surf1 proteins are facing each other, bound on opposite sides of each CIV core.

The presence of subunit extension and accessory subunits in CIV generates a pronounced cavity around the cytochrome *c*-binding site. However, overlaying of a *T. thermophila* cytochrome *c* homology model suggests that the canonical binding site is not obstructed. Cytochrome *c* binding is known to be driven by electrostatic interactions with the CuA domain of COX2, which forms a negatively charged patch in mammals. This structural feature is positively charged in *T. thermophila*, interacting mainly with H1 of cytochrome *c*, which displays a flipped polarity. We conclude that the experimentally observed functional incompatibility of *T. thermophila* CIV and mammalian cytochrome *c* is not due to divergent architecture, but to an inverted surface charge of the binding pocket.

Molecular dynamics simulations

We performed coarse-grained (CG) molecular dynamics simulations on the entire *T. thermophila* supercomplex structure using Martini3 forcefield[47] to study the rearrangement of the lipid bilayer around the highly bent protein assembly. Using the martinize2 (version 2.6) tool, we transformed the

atomistic structure into a CG model (atoms clustered into Martini beads)[48]. Using the small molecules database and the existing topologies of phospholipids available in the Martini3 forcefield[47], we generated the force-field parameters for cardiolipin. Cofactors and resolved lipids in the structure were not included in the simulated model system, and only protein was simulated to study the dynamics of lipid molecules around it. First, the CG model of the protein structure was minimized for 100 steps in vacuum to remove possible steric clashes. Then, the minimized CG supercomplex was embedded in a large (75 nm × 75 nm) hybrid membrane slab (POPE:POPC:CL in 4:2:1 ratio) using the insane.py script[49]. The CG protein–membrane system was solvated using standard Martini3 water beads and 100 mM Na^+ and Cl^- ions. Starting from this initial position, the simulation system was minimized keeping all beads free, first in double precision to resolve steric clashes between the lipids (maximum of 500 steps) and then in regular single precision (maximum of 10,000 steps). After minimization, with 4,000 kJ mol^{-1} nm^{-2} harmonic constraints on the backbone beads, the system was equilibrated using velocity-rescaling thermostat[50] and Berendsen barostat[51] for 10 ns. During production runs, the 4,000 kJ mol^{-1} nm^{-2} harmonic constraints on the backbone beads were maintained. Velocity-rescaling thermostat[50] and Parrinello–Rahman barostat[52] were used for temperature (310 K) and pressure (1 bar) control in the production phase. Coulombic interactions were treated with the reaction-field algorithm using $\varepsilon_r = 15$ (ref. [53]). The Verlet cut-off scheme was implemented with a Lennard–Jones cut-off of 1.1 nm (ref. [54]). The time step of the CG molecular dynamics simulations was 20 fs. Initial simulation replicas showed incomplete or unstable wrapping of the membrane around the protein, so we translated the lipid bilayer patch in *z*-direction and altered the insertion angle of the supercomplex to find an initial position that allowed the membrane to equilibrate and wrap fully around the protein (Systems T1-T7, simulation lengths of 0.9–2.8 µs, total of 9.6 µs). After finding the correct insertion of the protein into the membrane, we initiated three independent simulation replicas (Systems P1-P3, simulation lengths of 10 µs each, total of 30 µs). The simulations were performed using the Gromacs software (version 2021)[55]. In addition to molecular dynamics simulations of the entire supercomplex in the lipid bilayer, we also performed molecular dynamics simulations of the pure lipid bilayer (see

above; three replicas, 10 μs each) and of CIV_2, CI–$CIII_2$ and CI–$CIII_2$–CIV_2 subcomplexes (1–3 replicas, 1–5 μs each).

Data visualization and analysis

Images and videos were rendered using ChimeraX[56] and Visual Molecular Dynamics[57]. To analyse the *T. thermophila* cytochrome *c*-binding site, the mammalian cytochrome *c*-bound CIV structure (Protein Data Bank ID: 5IY5) was overlaid to the *T. thermophila* structure. Using AlphaFold2 (ref. [58]), a *T. thermophila* cytochrome *c* structure was predicted and overlaid to both the mammalian structure and the *T. thermophila* structures. The composite map of the complete respiratory supercomplex was generated in ChimeraX[56]. This map was only used for visualization, and not for atomic model refinement; instead, a consensus map was used. The buried areas of the CI–CII–$CIII_2$–CIV_2 supercomplex and the CIV_2 interface were calculated in ChimeraX[56].

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The atomic coordinates were deposited in the Protein Data Bank (PDB) under accession numbers 8BQS (supercomplex), 8B6F (CI), 8B6G (CII), 8B6H (CIV) and 8B6J (CIII). The cryo-EM maps have been deposited in the Electron Microscopy Data Bank (EMDB) under the respective accession numbers: EMD-16184, EMD-15865, EMD-15866, EMD-15867 and EMD-15868. The subtomogram averages have been deposited under EMD-15900. The atomic coordinates that were used in this study are: 1NTZ (cytochrome bc_1), 5IY5 (cytochrome *c*) and 5J4Z (ovine supercomplexes). The full versions of all gels are provided in the source file. An Excel file containing the visible absorption spectroscopy data and their analysis has been added. All the data will be publicly available. Source data are provided with this paper.

References

1. Schägger, H. & Pfeiffer, K. Supercomplexes in the respiratory chains of yeast and mammalian mitochondria. *EMBO J.* **19**, 1777–1783 (2000).

2. Caruana, N. J. & Stroud, D. A. The road to the structure of the mitochondrial respiratory chain supercomplex. *Biochem. Soc. Trans.* **48**, 621–629 (2020).

3. Bezawork-Geleta, A., Rohlena, J., Dong, L., Pacak, K. & Neuzil, J. Mitochondrial complex II: at the crossroads. *Trends Biochem. Sci.* **42**, 312–325 (2017).

4. Acin-Perez, R., Fernandez-Silva, P., Peleato, M. L., Perez-Martos, A. & Enriquez, J. A. Respiratory active mitochondrial supercomplexes. *Mol. Cell* **32**, 529–539 (2008).

5. Jiang, C. et al. Regulation of mitochondrial respiratory chain complex levels, organization, and function by arginyltransferase 1. *Front. Cell Dev. Biol.* **8**, 603688 (2020).

6. Lapuente-Brun, E. et al. Supercomplex assembly determines electron flux in the mitochondrial electron transport chain. *Science* **340**, 1567–1570 (2013).

7. Schon, E. A. & Dencher, N. A. Heavy breathing: energy conversion by mitochondrial respiratory supercomplexes. *Cell Metab.* **9**, 1–3 (2009).

8. Zhou, L., Maldonado, M., Padavannil, A., Guo, F. & Letts, J. A. Structures of *Tetrahymena*'s respiratory chain reveal the diversity of eukaryotic core metabolism. *Science* **376**, 831–839 (2022).

9. Colina-Tenorio, L., Horten, P., Pfanner, N. & Rampelt, H. Shaping the mitochondrial inner membrane in health and disease. *J. Intern. Med.* **287**, 645–664 (2020).

10. Blum, T. B., Hahn, A., Meier, T., Davies, K. M. & Kühlbrandt, W. Dimers of mitochondrial ATP synthase induce membrane curvature and

self-assemble into rows. *Proc. Natl Acad. Sci. USA* **116**, 4250–4255 (2019).

11. Mühleip, A., McComas, S. E. & Amunts, A. Structure of a mitochondrial ATP synthase with bound native cardiolipin. *eLife* **8**, e51179 (2019).

12. Kock Flygaard, R., Mühleip, A., Tobiasson, V. & Amunts, A. Type III ATP synthase is a symmetry-deviated dimer that induces membrane curvature through tetramerization. *Nat. Commun.* **11**, 5342 (2020).

13. Pinke, G., Zhou, L. & Sazanov, L. A. Cryo-EM structure of the entire mammalian F-type ATP synthase. *Nat. Struct. Mol. Biol.* **27**, 1077–1085 (2020).

14. Mühleip, A. et al. ATP synthase hexamer assemblies shape cristae of *Toxoplasma* mitochondria. *Nat. Commun.* **12**, 120 (2021).

15. Mühleip, A. W. et al. Helical arrays of U-shaped ATP synthase dimers form tubular cristae in ciliate mitochondria. *Proc. Natl Acad. Sci. USA* **113**, 8442–8447 (2016).

16. Allen, R. D., Schroeder, C. C. & Fok, A. K. An investigation of mitochondrial inner membranes by rapid-freeze deep-etch techniques. *J. Cell Biol.* **108**, 2233–2240 (1989).

17. Gu, J. et al. The architecture of the mammalian respirasome. *Nature* **537**, 639–643 (2016).

18. Guo, R., Zong, S., Wu, M., Gu, J. & Yang, M. Architecture of human mitochondrial respiratory megacomplex $I_2III_2IV_2$. *Cell* **170**, 1247–1257.e12 (2017).

19. Klusch, N., Senkler, J., Yildiz, O., Kuhlbrandt, W. & Braun, H. P. A ferredoxin bridge connects the two arms of plant mitochondrial complex I. *Plant Cell* **33**, 2072–2091 (2021).

20. Tobiasson, V. & Amunts, A. Ciliate mitoribosome illuminates evolutionary steps of mitochondrial translation. *eLife* **9**, e59264 (2020).

21. Tobiasson, V., Berzina, I. & Amunts, A. Structure of a mitochondrial ribosome with fragmented rRNA in complex with membrane-targeting elements. *Nat. Comm.* **13**, 6132 (2022).

22. Hagerhall, C. Succinate: quinone oxidoreductases. Variations on a conserved theme. *Biochim. Biophys. Acta* **1320**, 107–141 (1997).

23. Sun, F. et al. Crystal structure of mitochondrial respiratory membrane protein complex II. *Cell* **121**, 1043–1057 (2005).

24. Balabaskaran Nina, P. et al. Highly divergent mitochondrial ATP synthase complexes in *Tetrahymena thermophila*. *PLoS Biol.* **8**, e1000418 (2010).

25. Berndtsson, J. et al. Respiratory supercomplexes enhance electron transport by decreasing cytochrome *c* diffusion distance. *EMBO Rep.* **21**, e51015 (2020).

26. Tremmel, I. G., Kirchhoff, H., Weis, E. & Farquhar, G. D. Dependence of plastoquinol diffusion on the shape, size, and density of integral thylakoid proteins. *Biochim. Biophys. Acta* **1607**, 97–109 (2003).

27. Kirchhoff, H. Diffusion of molecules and macromolecules in thylakoid membranes. *Biochim. Biophys. Acta* **1837**, 495–502 (2014).

28. Letts, J. A., Fiedorczuk, K. & Sazanov, L. A. The architecture of respiratory supercomplexes. *Nature* **537**, 644–648 (2016).

29. Althoff, T., Mills, D. J., Popot, J. L. & Kuhlbrandt, W. Arrangement of electron transport chain components in bovine mitochondrial supercomplex $I_1III_2IV_1$. *EMBO J.* **30**, 4652–4664 (2011).

30. Wu, M., Gu, J., Guo, R., Huang, Y. & Yang, M. Structure of mammalian respiratory supercomplex $I_1III_2IV_1$. *Cell* **167**, 1598–1609.e10 (2016).

31. Gao, X. et al. Structural basis for the quinone reduction in the bc_1 complex: a comparative analysis of crystal structures of mitochondrial

cytochrome bc_1 with bound substrate and inhibitors at the Q_i site. *Biochemistry* **42**, 9067–9080 (2003).

32. Zhang, Z. et al. Electron transfer by domain movement in cytochrome bc_1. *Nature* **392**, 677–684 (1998).

33. Rajagukguk, S. et al. Effect of mutations in the cytochrome *b* ef loop on the electron-transfer reactions of the Rieske iron–sulfur protein in the cytochrome bc_1 complex. *Biochemistry* **46**, 1791–1798 (2007).

34. Chen, M. et al. Distinct structural modulation of photosystem I and lipid environment stabilizes its tetrameric assembly. *Nat. Plants* **6**, 314–320 (2020).

35. Wolf, D. M. et al. Individual cristae within the same mitochondrion display different membrane potentials and are functionally independent. *EMBO J.* **38**, e101056 (2019).

36. Davies, K. M. et al. Macromolecular organization of ATP synthase and complex I in whole mitochondria. *Proc. Natl Acad. Sci. USA* **108**, 14121–14126 (2011).

37. Paumard, P. et al. The ATP synthase is involved in generating mitochondrial cristae morphology. *EMBO J.* **21**, 221–230 (2002).

38. Jha, P., Wang, X. & Auwerx, J. Analysis of mitochondrial respiratory chain supercomplexes using blue native polyacrylamide gel electrophoresis (BN-PAGE). *Curr. Protoc. Mouse Biol.* **6**, 1–14 (2016).

39. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018).

40. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).

41. Hagen, W. J. H., Wan, W. & Briggs, J. A. G. Implementation of a cryo-electron tomography tilt-scheme optimized for high resolution

subtomogram averaging. *J. Struct. Biol.* **197**, 191–198 (2017).

42. Kremer, J. R., Mastronarde, D. N. & McIntosh, J. R. Computer visualization of three-dimensional image data using IMOD. *J. Struct. Biol.* **116**, 71–76 (1996).

43. Nicastro, D. et al. The molecular architecture of axonemes revealed by cryoelectron tomography. *Science* **313**, 944–948 (2006).

44. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).

45. Afonine, P. V. et al. Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. D* **74**, 531–544 (2018).

46. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).

47. Souza, P. C. T. et al. Martini 3: a general purpose force field for coarse-grained molecular dynamics. *Nat. Methods* **18**, 382–388 (2021).

48. de Jong, D. H. et al. Improved parameters for the Martini coarse-grained protein force field. *J. Chem. Theory Comput.* **9**, 687–697 (2013).

49. Wassenaar, T. A., Ingolfsson, H. I., Bockmann, R. A., Tieleman, D. P. & Marrink, S. J. Computational lipidomics with insane: a versatile tool for generating custom membranes for molecular simulations. *J. Chem. Theory Comput.* **11**, 2144–2155 (2015).

50. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101 (2007).

51. Berendsen, H. J. C., Postma, J. P. M., Vangunsteren, W. F., Dinola, A. & Haak, J. R. Molecular-dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690 (1984).

52. Parrinello, M. & Rahman, A. Polymorphic transitions in single-crystals—a new molecular-dynamics method. *J. Appl. Phys.* **52**, 7182–7190

(1981).

53. Tironi, I. G., Sperb, R., Smith, P. E. & Vangunsteren, W. F. A generalized reaction field method for molecular-dynamics simulations. *J. Chem. Phys.* **102**, 5451–5459 (1995).

54. Grubmuller, H., Heller, H., Windemuth, A. & Schulten, K. Generalized Verlet algorithm for efficient molecular dynamics simulations with long-range interactions. *Mol. Simul.* **6**, 121–142 (1991).

55. Abraham, M. J. et al. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).

56. Goddard, T. D. et al. UCSF ChimeraX: meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).

57. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graphics* **14**, 33–38 (1996).

58. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

Acknowledgements

We thank SciLifeLab EM facility (funded by the KAW, EPS and Kempe foundations), the Center for Scientific Computing, Finland for high-performance supercomputing resources, the EMBO Young Investigator Program, the Swedish Foundation for Strategic Research (FFL15:0325), the Ragnar Söderberg Foundation (M44/16), Cancerfonden (2017/1041), the European Research Council (ERC-2018-StG-805230) and the Knut and Alice Wallenberg Foundation (2018.0080). V.S. was supported by the Academy of Finland, the Sigrid Jusélius Foundation, the Jane and Aatos Erkko Foundation, the Magnus Ehrnrooth Foundation and the University of Helsinki. A.Maréchal was supported by the Medical Research Council (CDA MR/M00936X/1, transition support MR/T032154/1). We also thank M. Kozlov, B. Bruininks, W. Kulig, P. C. T. Souza and M. Girych for discussions on molecular dynamics simulations.

Funding

Open access funding provided by Stockholm University.

Author information

Author notes

1. Alexander Mühleip

   Present address: School of Infection and Immunity, University of Glasgow, Wellcome Centre for Integrative Parasitology, Glasgow, UK

2. Rasmus Kock Flygaard

   Present address: Department of Molecular Biology and Genetics, Danish Research Institute of Translational Neuroscience-DANDRITE, Nordic EMBL Partnership for Molecular Medicine, Aarhus University, Aarhus C, Denmark

3. Rozbeh Baradaran & Victor Tobiasson

   Present address: MRC Laboratory of Molecular Biology, Cambridge, UK

4. Victor Tobiasson

   Present address: National Center for Biotechnology Information, National Library of Medicine, National Institute of Health, Bethesda, MD, USA

5. These authors contributed equally: Alexander Mühleip, Rasmus Kock Flygaard

Authors and Affiliations

1. Science for Life Laboratory, Department of Biochemistry and Biophysics, Stockholm University, Solna, Sweden

Alexander Mühleip, Rasmus Kock Flygaard, Rozbeh Baradaran, Victor Tobiasson & Alexey Amunts

2. Department of Physics, University of Helsinki, Helsinki, Finland

Outi Haapanen & Vivek Sharma

3. Institute of Structural and Molecular Biology, Birkbeck College, London, UK

Thomas Gruhl & Amandine Maréchal

4. Institute of Structural and Molecular Biology, University College London, London, UK

Amandine Maréchal

5. HiLIFE Institute of Biotechnology, University of Helsinki, Helsinki, Finland

Vivek Sharma

Contributions

A. Mühleip, R.K.F. and A.A. designed the project. V.T. performed cell culturing and isolation of mitochondria. A. Mühleip and R.K.F. prepared the sample and collected cryo-EM data. A. Mühleip, R.K.F. and R.B. processed cryo-EM data and built the model. A.Mühleip performed cryo-electron tomography and subtomogram averaging. O.H. and V.S. performed molecular dynamics simulations. T.G., A. Maréchal and A.A. performed biochemical and spectroscopic analyses. A. Mühleip, R.K.F. and A.A. wrote the manuscript with contributions from O.H., V.S. and A.Maréchal. All authors contributed to revising the manuscript.

Corresponding author

Correspondence to Alexey Amunts.

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks Adam Frost and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Cryo-EM and subtomogram averaging.

(A) A total of 26,063 movies were recorded and analyzed, a representative micrograph is shown. Particles that were not respiratory supercomplex were discarded by classification, since they cannot contribute to reconstruction. Cryo-EM structures were successfully obtained from three preliminary datasets. **(B)** 2D class averages. **(C)** Fourier Shell Correlation according to the 0.143 gold-standard criterion. **(D)** processing workflow in RELION-3.1 (R) and cryoSPARC2 (C). **(E)** Fourier Shell correlation of the subtomogram average indicating a resolution of 28 Å. **(F)** Subtomogram average map-model overlay.

Extended Data Fig. 2 Bound native lipids of the ciliate I-II-III$_2$-IV$_2$ supercomplex reveal a curved membrane region.

Top view **(A)** and side view **(B)** with bound lipids. Cardiolipin (red), phosphatidylcholine (blue) phosphatidylethanolamine (green), phosphatidic acid (yellow), ubiquinone-8 (purple). **(C-F)** Location of the four ubiquinone-8 molecules with closeup views **(D-F)**. **(G)** Coomassie-stained CN-PAGE and in-gel activity assay to visualize the purified I-II-III$_2$-IV$_2$ supercomplex. Individual lanes highlight protein bands with active CI, CII (purple) and CIV (brown), respectively. The band marked with an asterisk (*) was tentatively assigned as the intact supercomplex.

Extended Data Fig. 3 Lineage-specific structural features of CI dictate a divergent supercomplex architecture.

(**A**) Canonical subunits of CI (coloured) include conserved antiporter-like folds (ND2a/b, ND4, ND5a/b), and γ-carbonic anhydrase trimer. **(B)** Augmented CI architecture occluding canonical CIV binding site. **(C)** The ferrodoxin bridge of NDUAB1-a and NDUFX connects ND2a of the membrane arm and NDUA6 of the matrix arm. **(D)** The matrix arm contains bound subunits NDUTT1 with UDP and NDUTT3 with ADP. **(E)** NDUFA9 contains a bound NADPH molecule close to the CI Q-tunnel. **(F)** Redox centers in CI matrix arm. **(G)** zinc ion coordinated by NDUS6.

Extended Data Fig. 4 CIV$_2$ contains a bound Tim hexamer and cytochrome-*c* binding site with inverted electrostatic charge.

(A) sideview of CIV$_2$, containing numerous accessory subunits, including a Tim hexamer. **(B)** Closeup view of COX2, which contains an insertion that recruits the Tim hexamer to the CIV dimer. Furthermore, COX2 interacts with the two Surf1-like proteins COXTT8 and COXTT4. **(C-D)** ciliate and mammalian CIV$_2$ structures. An overlay of a predicted Tt-cytochrome-c structure fits the cytochrome-c binding crater without clashes.

Extended Data Fig. 5 The augmented CIV_2 contains numerous associated compact-fold subunits.

(A,B) Matrix and lumen view of the CIV dimer, subunit locations of insets are indicated. (**C,D**) Subunits COXMC1, MC2 and MC3 form mitochondrial carrier folds. COXFS is a CDGSH iron sulfur domain **(E). (F)** Ymf75 coordinates a noncanonical heme group at the CIV periphery. (**G-K**) Compact folds of accessory subunits include a seven-bladed beta-propeller (G), two Surf1-like proteins (H, I), a TraB-family protein (J) and a Pleckstrin homology (PH) like domain (K).

Extended Data Fig. 6 Fragmentation and extension of COX3 and CI proton-pump subunits enables formation of a divergent supercomplex assembly.

(A) The supercomplex displays additional subunits and ordered structural elements (colored cartoons) compared to the amphipol-embedded CIV_2 dimer (PDB 7W5Z). In the CI-IV interface this includes subunit NDUTT16 and a globin-like domain of COXTT2. **(B)** Matrix and lumenal views of the CIV_2 interface include NDUFA4. **(C)** The CIV_2 contains 210 lipids, many of which populate the dimer interface. **(D)** The *T. thermophila* dimer is augmented compared to the bovine dimer and contains the previously unassigned canonical subunit NDUFA4, which is absent in the bovine dimer (right, PDB 3X2Q). Vectors (red, blue) indicate different distances between COX1 centers. (**E)** Overlay of the mammalian COX3 (PDB 5IY5) and the Tt-COX3 (this study) shows that the N-terminal fragment (COX3a), corresponds to Helix-1 (H1) of the conserved structure. COX3b makes up most of the conserved COX3 fold and contains large extensions that contribute to supercomplex formation. **(F)** *T. thermophila* mitochondrial genome region showing the insertion of the tRNA-Trp gene in between *ymf67*/COX3a and *ymf68*/COX3b (genes in green, protein/RNA transcripts in red and purple). **(G)** ciliate CI outline (blue) with the antiporter-like subunits shown as cylinders. The canonical proton pumps ND1, ND2 and ND5 are encoded by split genes, resulting in chains *a* and *b*. **(H)** mammalian CI (blue, PDB 5LNK) with single-chain ND1-6 subunits color-coded as in G.

Extended Data Fig. 7 Conservation and divergence of ciliate CII.

(A) Ciliate CII (left) colored by subunits and shown in comparison to mammalian CII (right, PDB 1ZOY) with similar color scheme for subunits SDHA-D. SDHC and SDHD are substantially reduced and partially replaced by ciliate-specific subunits. Dashed box highlights a noncanonical cysteine-linked heme C that is coordinated by two axial histidines in SDHTT1. **(B)** Conserved electron transfer chain in CII. **(C)** CII colored by conserved subunits (grey), extensions (gold) and ciliate-specific subunits (red). **(D)** Individual CII subunits with similar color coding as in C. **(E)** CII contains a conserved proximal ubiquinone.

Extended Data Fig. 8 UV-visible redox absorption spectrum indicate activity of the supercomplex.

Dithionite minus air-oxidized difference spectrum. Peaks are labeled that correspond to the absorption of the A-type hemes present in CIV (449/610 nm, aa_3) as well as merged absorption bands of other B- and C-type hemes (maxima at 432/558 nm). Inset: Deconvolution of the 558-nm absorption band highlighting the contribution of the B- and C-type hemes of CIII (b_H at 561 nm, purple; b_L at 558 and 565 nm, blue; c_1 at 552 nm, green) and the presence of at least another heme-protein with absorption maximum at 556 nm (orange). The spectrum shown is representative of two separate supercomplex preparations.

Source data

Extended Data Fig. 9 Extensive CI-CIII$_2$ interface involves numerous lipids and accommodates flexible head domain of distal UQCRFS1.

(A) Buried surfaces in the CI-CIII$_2$ interface. CIII$_2$ contacts on CI (orange) and CI contact points on CIII$_2$ (blue; >9,000 Å^2 interface). **(B)** Four Tt-CI subunits would clash with mammalian CIII$_2$ in canonical position, thus Tt-

$CIII_2$ is displaced 14 Å from Tt-CI. **(C)** Tt-CI (light blue) and Tt-$CIII_2$ (light orange) with interface-forming subunits shown in colored ribbons. **(D)** Orthogonal views of bound native lipids in the I-III_2 interface. **(E)** Distances between ciliate (grey) and mammalian (purple) heme b_H, heme b_L and heme c_1 groups (measuring from Fe atoms) and the ciliate Q_o sites. Superposing mammalian $CIII_2$ (PDB 7O3H) shows that ciliate heme b_H and heme b_L are displaced further away from the symmetry-related hemes, whereas the ciliate heme c_1 groups are slightly closer. Furthermore, the distance from Q_o sites (dark gold) to heme b_L is comparable to bovine $CIII_2$ (17). **(F)** The $CIII_2$ wedged ubiquinone (UQ, transparent gold density) in COB (green) is almost rotationally symmetric, showing planar head group density equally close to the heme b_L molecules (transparent blue). **(G)** Cryo-EM density carved around the UQCRFS1 proteins in $CIII_2$. Head domains show inherent flexibility. **(H)** Location of distal and proximal UQCRFS1 subunits (yellow ribbon) with respect to CI quinone tunnel (black box). The purple box highlights region for focused 3D classification. **(I)** 3D classes I-VIII. UQCRFS1 and UQCRTT3 were fitted to each map and distance measured from Fe_2S_2 cluster to heme c_1. Only classes V and VI display markedly different distances. **(J)** Comparison of B-state class V (left) and C-state class VI (right). B-state UQCRFS1 superposes well with B-state UQCRFS1 from chicken $CIII_2$ (purple, PDB 3BCC). C-state UQCRFS1 superposes almost perfectly with C-state UQCRFS1 from chicken $CIII_2$ (purple, PDB 1BCC). Measured distances are from UQCRFS1 Fe_2S_2 to heme c_1 and b_L.

Extended Data Fig. 10 Membrane curvature induction by the supercomplex and subcomplexes.

(A) Left, selected frames of molecular dynamics simulations of the I-II-III_2-IV_2 supercomplex and CIV_2, embedded in a lipid bilayer. Contact sites of protein complexes with sliced plane are shown by black markers. Right, overlay and zoom in of the frames of the I-II-III_2-IV_2 supercomplex and CIV_2. The larger size of the supercomplex and its architecture bends the membrane (blue marker) more than CIV_2 alone (red marker). **(B)** Detergent/saponin belts extracted from the cryo-EM structures of CIV_2

(PDB 7W5Z), CI-$CIII_2$ (PDB 7TGH) and of the I-II-III_2-IV_2 supercomplex. Lengths of curvature arcs when viewed along the membrane tube are indicated. **(C)** Selected frame of the molecular dynamics simulation of the I-III_2-IV_2 supercomplex (missing CII), displaying a similar curvature to the full supercomplex.

Supplementary information

Supplementary Information

This file contains Supplementary Table 1: Data collection and model statistics; Supplementary Table 2: List of proteins and comments; and Source Data for Extended Data Fig. 2.

Reporting Summary

Peer Review File

Supplementary Video 1

Coarse-grained molecular dynamics simulation of the *T. thermophila* supercomplex. First 800 ns of the molecular dynamics simulation starting from an initially planar membrane reveals a deformation of the bilayer into a curved topology to accommodate the membrane protein complex.

Supplementary Video 2

Coarse-grained molecular dynamics simulation of pure lipid bilayer. Membrane shows fluctuations, however, a stable curved architecture (as in the case of supercomplex) is not observed.

Supplementary Video 3

Annular lipid shell of the *T. thermophila* supercomplex. Final frame of the coarse-grained molecular dynamics simulation with supercomplex and

surrounding annular lipids shown, highlighting the curved shape of the lipid belt.

Source data

Source Data Extended Data Fig. 8

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Mühleip, A., Flygaard, R.K., Baradaran, R. *et al.* Structural basis of mitochondrial membrane bending by the I–II–III_2–IV_2 supercomplex. *Nature* **615**, 934–938 (2023). https://doi.org/10.1038/s41586-023-05817-y

- Received: 28 June 2022
- Accepted: 09 February 2023

- Published: 22 March 2023
- Issue Date: 30 March 2023
- DOI: https://doi.org/10.1038/s41586-023-05817-y

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

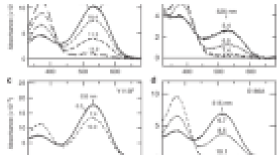
Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

---

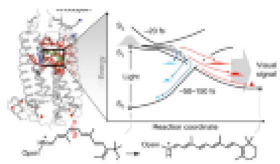
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05817-y



The counterion–retinylidene Schiff base interaction of an invertebrate rhodopsin rearranges upon light activation

13 May 2019

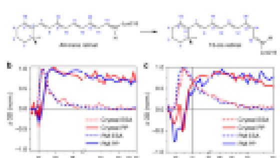
Takashi Nagata, Mitsumasa Koyanagi, … Akihisa Terakita



Quantum–classical simulations of rhodopsin reveal excited-state population splitting and its effects on quantum efficiency

03 March 2022

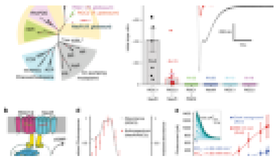
Xuchun Yang, Madushanka Manathunga, … Massimo Olivucci



Three-dimensional view of ultrafast dynamics in photoexcited bacteriorhodopsin

18 July 2019

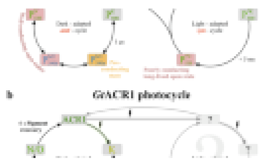
Gabriela Nass Kovacs, Jacques-Philippe Colletier, … Ilme Schlichting



NeoR, a near-infrared absorbing rhodopsin

10 November 2020

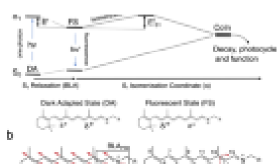
Matthias Broser, Anika Spreen, … Peter Hegemann



Time-resolved spectroscopic and electrophysiological data reveal insights in the gating mechanism of anion channelrhodopsin

14 May 2021

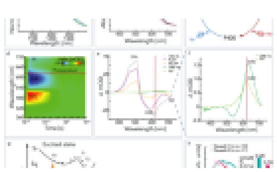
Max-Aylmer Dreier, Philipp Althoff, … Klaus Gerwert



Retinal chromophore charge delocalization and confinement explain the extreme photophysics of Neorhodopsin

04 November 2022

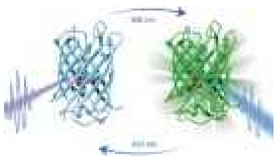
Riccardo Palombo, Leonardo Barneschi, … Massimo Olivucci



[QuasAr Odyssey: the origin of fluorescence and its voltage sensitivity in microbial rhodopsins]

20 September 2022

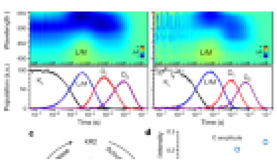
Arita Silapetere, Songhwan Hwang, … Peter Hegemann



[Infrared spectroscopy reveals multi-step multi-timescale photoactivation in the photoconvertible protein archetype dronpa]

11 June 2018

Sergey P. Laptenok, Agnieszka A. Gil, … Stephen R. Meech



[Femtosecond-to-millisecond structural changes in a light-driven sodium pump]

20 May 2020

Petr Skopintsev, David Ehrenberg, … Jörg Standfuss

- Article
- Open Access
- [Published: 22 March 2023]

Ultrafast structural changes direct the first molecular events of vision

- Thomas Gruhl,
- Tobias Weinert,
- Matthew J. Rodrigues,
- Christopher J. Milne,
- Giorgia Ortolani,
- Karol Nass,
- Eriko Nango,
- Saumik Sen,
- Philip J. M. Johnson,
- Claudio Cirelli,
- Antonia Furrer,
- Sandra Mous,
- Petr Skopintsev,
- Daniel James,
- Florian Dworkowski,
- Petra Båth,
- Demet Kekilli,
- Dmitry Ozerov,
- Rie Tanaka,
- Hannah Glover,
- Camila Bacellar,
- Steffen Brünle,
- Cecilia M. Casadei,
- Azeglio D. Diethelm,
- Dardan Gashi,
- Guillaume Gotthard,
- Ramon Guixà-González,
- Yasumasa Joti,
- Victoria Kabanova,
- Gregor Knopp,
- Elena Lesca,
- Pikyee Ma,

- Isabelle Martiel,
- Jonas Mühle,
- Shigeki Owada,
- Filip Pamula,
- Daniel Sarabi,
- Oliver Tejero,
- Ching-Ju Tsai,
- Niranjan Varma,
- Anna Wach,
- Sébastien Boutet,
- Kensuke Tono,
- Przemyslaw Nogly,
- Xavier Deupi,
- So Iwata,
- Richard Neutze,
- Jörg Standfuss,
- Gebhard Schertler &
- …
- Valerie Panneels

Show authors

Nature **volume 615**, pages 939–944 (2023)

- 16k Accesses
- 1 Citations
- 353 Altmetric
- Metrics details

Abstract

Vision is initiated by the rhodopsin family of light-sensitive G protein-coupled receptors (GPCRs)[1]. A photon is absorbed by the 11-*cis* retinal chromophore of rhodopsin, which isomerizes within 200 femtoseconds to

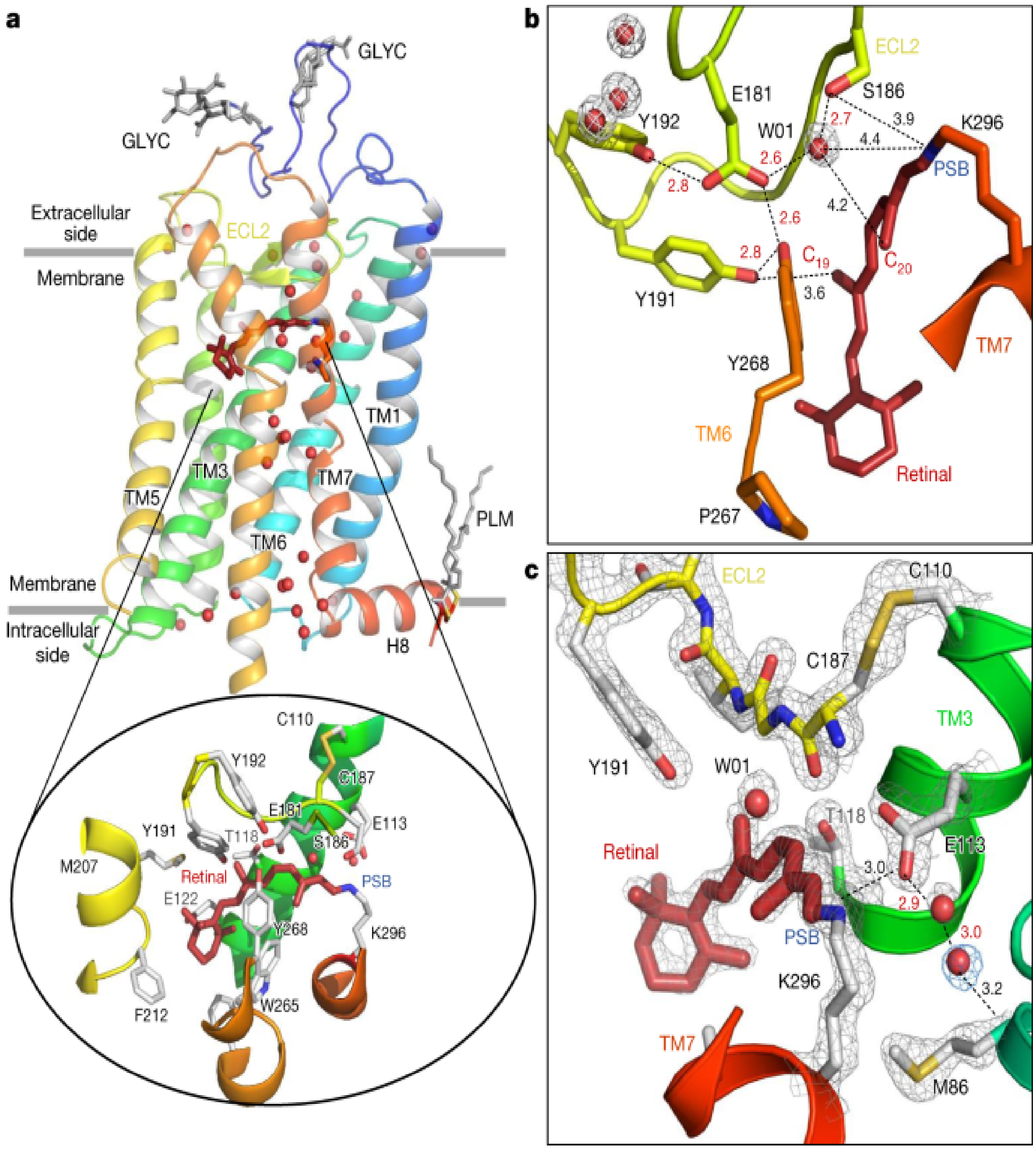
the all-*trans* conformation[2], thereby initiating the cellular signal transduction processes that ultimately lead to vision. However, the intramolecular mechanism by which the photoactivated retinal induces the activation events inside rhodopsin remains experimentally unclear. Here we use ultrafast time-resolved crystallography at room temperature[3] to determine how an isomerized twisted all-*trans* retinal stores the photon energy that is required to initiate the protein conformational changes associated with the formation of the G protein-binding signalling state. The distorted retinal at a 1-ps time delay after photoactivation has pulled away from half of its numerous interactions with its binding pocket, and the excess of the photon energy is released through an anisotropic protein breathing motion in the direction of the extracellular space. Notably, the very early structural motions in the protein side chains of rhodopsin appear in regions that are involved in later stages of the conserved class A GPCR activation mechanism. Our study sheds light on the earliest stages of vision in vertebrates and points to fundamental aspects of the molecular mechanisms of agonist-mediated GPCR activation.

Main

Rhodopsin, the vertebrate receptor for low-light vision, is concentrated within the disk membranes of rod cells in the retina. Rhodopsin transforms the absorption of light into a physiological signal through conformational changes that activate the intracellular G protein transducin—a member of the Gi/o/t family—initiating a signalling cascade, resulting in electrical impulses sent to the brain and ultimately leading to visual perception. The structure of rhodopsin consists of seven transmembrane (TM) α-helices with an 11-*cis* retinal chromophore covalently bound through a protonated Schiff base (PSB) to Lys296$^{7.43}$ of TM7 (the superscript values on amino acids containing the TM domain location refer to the Ballesteros–Weinstein scheme, explained in the ‘Residue numbering’ section of the Methods). This buried ligand is located within the TM bundle towards the extracellular side, like in many class A GPCRs. Retinal also contacts extracellular loop 2 (ECL2), which forms a lid over the chromophore and contains a highly conserved disulfide bridge (Cys110$^{3.25}$–Cys187^{ECL2}) connecting to the central helix TM3 (see box 1 of ref. [1]). Glu113$^{3.28}$ provides a negatively

charged counterion[4] that forms a salt bridge with the PSB (Fig. 1a–c) and thereby participates in the stabilization of the receptor resting state[5]. From our understanding of the evolution of visual pigments[6,7], we know that, originally, Glu181^{ECL2} was the only residue able to neutralize the positive charge of the Schiff base. This 'ancestral counterion'[8], which still functions as a complex counterion in invertebrates[6], remains connected through a water-mediated hydrogen bond network to the PSB of vertebrate rhodopsins (Fig. 1b). The second, main counterion Glu113$^{3.28}$ appeared during evolution and both residues are important for the activation mechanism. Structures of light-activated rhodopsin trapped at low temperature[9,10], structures of the late Meta II active state[11,12,13], and copious computational[14], biochemical and spectroscopic studies[15,16,17] have provided important insights into the mechanism of signal transduction in rhodopsin. However, methods that provide both a high spatial and temporal resolution are required to obtain a complete experimentally derived picture of the activation mechanism at the atomic scale from femtoseconds to milliseconds.

Fig. 1: Room temperature SFX structure of the dark state of bovine rhodopsin from crystals grown in LCP.



a, The overall structure of rhodopsin, rainbow coloured by residue number from blue (N terminus) to red (C terminus). The seven-TM bundle contains two *N*-glycosylation domains (GLYC) and palmitate groups (PLM) that anchor the amphipathic helix H8 to the membrane (grey lines). Water molecules (red spheres) form key networks[1] between the extracellular (retinal ligand-binding pocket) and intracellular (G protein-binding site) regions of the receptor. The 11-*cis* retinal (dark red) is covalently bound to Lys296 (inset) through the PSB. The retinal-binding pocket is further

composed of amino acids surrounding the PSB (the counterion Glu113, and Met44, Phe91, Thr94, Ala292 and Phe293), the retinal aliphatic chain (Ala117, Thr118, Tyr191, Trp265 and Glu181/Ser186 through water W01) and the β-ionone ring (Gly120, Gly121, Glu122, Phe212, Met207, Phe261 and Ala269; for clarity, only selected residues in the binding pocket are shown). **b**,**c**, Examples of well-resolved molecules in the water-mediated networks connecting the residues in the ancestral counterion Glu181 network (**b**) and the counterion Glu113 to Met86 of TM2 (and Ala117$^{3.32}$, not shown) (**c**). The water molecules have well-defined electron densities (grey and blue meshes, $2F_{obs} - F_{calc}$ electron density contoured at 2.2 and 0.7σ, respectively).

In recent years, time-resolved crystallography[3,18] at X-ray free-electron lasers (XFEL) has been used to reveal ultrafast structural changes in myoglobin[19], photoactive yellow protein[20,21], fluorescent proteins[22], bacterial phytochromes[23], microbial proton[24,25], sodium[26] and chloride[27] pumps, a bacterial photosynthetic reaction centre[28] and also with non-photosensory proteins with exogenously added non-natural photocaged ligands[29,30]. In time-resolved serial femtosecond crystallography (TR-SFX), the protein molecules in the crystals are photoactivated with an optical laser pulse and the structure is probed with an X-ray pulse from an XFEL after a specified time delay. As each crystal generates one diffraction pattern, the experiment is carried out in a serial manner: the frames are collected from tens of thousands of randomly oriented crystals. The TR-SFX method is complementary to spectroscopy methods, revealing structural detail at the atomic level in the femtosecond domain, without directly resolving charge effects, hydrogen bond interactions and electronic changes. The atomic resolution comes at the cost of less clearly defined illumination conditions[31], which are a matter of discussion[24,25,32], and systematic studies investigating these in more detail are underway[33].

Here we used TR-SFX at room temperature to follow the light-induced conversion of the inverse agonist 11-*cis* retinal to an agonist all-*trans* in a vertebrate opsin. Our observations reveal how this translates into early structural changes within the protein. After 1 ps, we observe a twisted retinal that stores energy while structural motions in the protein radiate as an

anisotropic propagation away from the retinal chromophore. The rhodopsin structure, 100 ps later, reveals a slightly more relaxed conformation.

Room temperature structure of rhodopsin

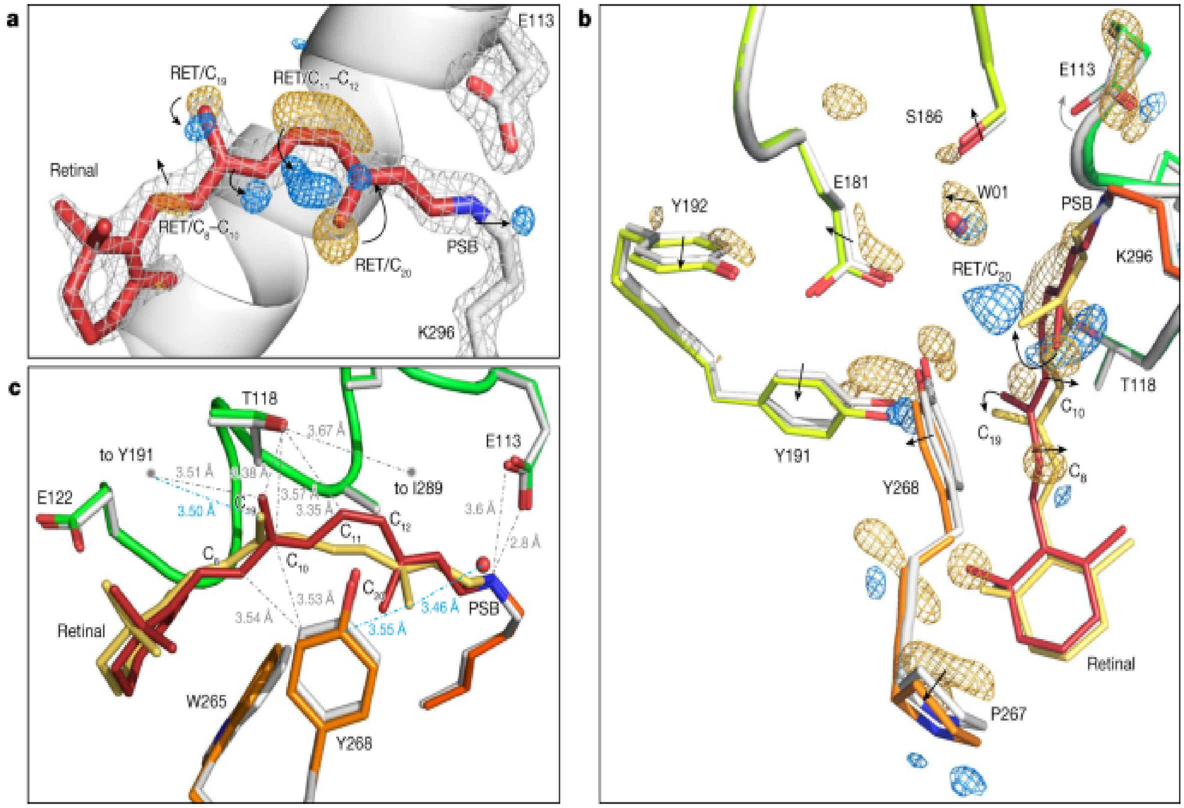
The room-temperature SFX structure of rhodopsin in the inactive dark state was obtained at a resolution of 1.8 Å (Extended Data Table 1 (dark state)) from microcrystals grown in a lipidic cubic phase (LCP). As with most membrane protein structures determined from LCP-grown crystals, these crystals display a type I lattice[34] forming stacks of protein two-dimensional layers built through hydrophobic interactions (Extended Data Fig. 1a–c). Potentially physiologically relevant dimers of rhodopsin molecules[35] form contacts between the TM1 and helix 8 (H8) segments of each monomer and are assembled in a head-to-tail manner generating the asymmetric unit. By close inspection of the diffraction data and the resulting electron density maps, the presence of translation-related crystal domains was detected. Measured intensities were corrected to account for this, globally improving the quality and interpretability of the maps[36] (Methods, Extended Data Fig. 1d–i and Extended Data Table 1). Overall, the SFX structure in the inactive dark state of rhodopsin (Fig. 1a) is very similar to other crystal structures collected at cryogenic temperatures (for example, Protein Data Bank (PDB) 1GZM; root mean square deviation = 0.33 Å on C_α atoms)[37]. In contrast to earlier structures solved in cryogenic conditions, the present room temperature structure reveals electron density for all of the previously described functional and structural water molecules. These include the water-mediated cluster around the ancestral counterion Glu181 and its polar tyrosine cage (Fig. 1b), which have a role later in the photoactivation process[38,39]. Moreover, a new ordered water molecule was resolved near the Schiff base connecting the proximal counterion Glu113$^{3.28}$ to Met86$^{2.53}$ (Fig. 1c) and Ala117$^{3.32}$. This interaction has a central position at the TM2–TM3–TM7 interface in the TM bundle.

A picosecond light-induced bent retinal

The first metastable intermediate of rhodopsin (bathorhodopsin, Batho-Rh)[2,15,40] arises 200 fs after photoactivation. It is fully populated by $\Delta t = 1$ ps

(refs. [14,41]) and persists for tens of nanoseconds[42,43]. To characterize the structure of Batho-Rh, we collected TR-SFX data at the Swiss and Japanese XFELs (Extended Data Fig. 2) from LCP-grown microcrystals of rhodopsin photoactivated under a regime in which laser-induced heating is low (Extended Data Fig. 3, Methods) and in which we recovered high-quality difference electron density maps below the maximum activation (Extended Data Fig. 3) using a femtosecond-pump laser with a 480-nm wavelength for three time delays of photoactivation of $\Delta t = 1$ ps, 10 ps and 100 ps. High-quality TR-SFX data (Extended Data Table 1) are represented as difference Fourier electron density maps in Fig. 2a,b and Extended Data Fig. 4. For the shorter time delay, changes in electron density are distributed in a highly anisotropic manner, clustering in the immediate vicinity of the buried retinal chromophore and propagating towards the cytoplasmic side of the protein through the TM5 and TM6 helices (Extended Data Fig. 5) at a minimum speed of 18 Å ps^{-1} measured along TM6 (1,800 m s^{-1}, slightly above the speed of sound in water and in accordance with the speed of sound in ribonuclease A crystals[44]). This structural anisotropy had completely decayed by $\Delta t = 100$ ps.

Fig. 2: Retinal conformation captured 1 ps after rhodopsin photoactivation using TR-SFX.



a, Changes in electron density. The retinal (RET) model (red) and the contoured grey mesh (at 2.7σ of the $2F_{obs} - F_{calc}$ electron density map) correspond to rhodopsin in the dark state obtained by SFX. The difference Fourier electron density (F_{obs}(light) $-$ F_{obs}(dark) contoured at 3.8σ) around the C_{11}=C_{12} bond of the polyene chain and the C_{20} methyl show features appearing after 1 ps photoactivation in blue (positive density) that are correlated with disappearing features in gold (negative density), establishing that the chromophore has already isomerized. A negative density is also observed along C_8 and C_{10} of the retinal polyene chain. **b**,**c**, The effect of retinal isomerization on the surrounding amino acid residues. The model of 1-ps-photoactivated rhodopsin (retinal in yellow; rhodopsin in orange and green) obtained from the extrapolated map $2F_{ext} - F_{calc}$ (21% photoactivation; Methods) superimposed to the dark-state model (retinal in red; rhodopsin in grey). The main chain C_α atoms of the protein were used for the structural superposition. **b**, The difference electron density map (F_{obs}(light) $-$ F_{obs}(dark), contoured at 3.4σ) shows the presence of positive and negative electron densities (blue and yellow) around specific amino acids such as Tyr268[6.51] of the binding pocket. The arrows illustrate shifts or

rotations. **c**, The torsion of the retinal polyene chain at C_{11}–C_{13} in the direction of Tyr268 (the π-system at the isomerizing bond of retinal (yellow model) is now rotated 90° with respect to that of the dark state[48] (red model) (Extended Data Fig. 6)) and the bending along C_6–C_{11}. Selected distances from retinal to rhodopsin residues are shown as grey dotted lines for the dark state and as blue dotted lines for the isomerized form.

Light-induced structural changes within the retinal polyene chain are observed at $\Delta t = 1$ ps as a strong negative difference electron density feature (minimum of -6.2σ, where σ is the root mean square electron density of the unit cell) and a complementary positive difference electron density feature (maximum of $+5.8\sigma$) (Fig. 2a) associated with the C_{11}=C_{12} double bond revealing that this bond has isomerized. This event is associated with changes in electron density near the C_{20} methyl (-6.6σ and $+5.6\sigma$). Modelling these electron density changes in combination with structural refinement against crystallography observations extrapolated to 100% occupancy of the photoactivated intermediate (for photoactivation levels, see the Methods and Extended Data Table 1; 1-, 10- and 100-ps time delays) establishes that retinal isomerization involves a large (47.7°) clockwise rotation (as viewed from the PSB) of the C_{20} methyl (or anticlockwise if adopting the conventional dihedral angle sign with view from the β-ionone side). This rotation happens towards the extracellular side in concert with a shift in both the proximal water molecule W01 and Tyr268$^{6.51}$ of the Glu181 cage (Fig. 2b). The tilt of the C_{20} methyl increases (Extended Data Fig. 6e) to 51.3° by $\Delta t = 100$ ps. A similar distortion was measured as 54.9° in a cryo-trapped Batho-Rh state study that lacked temporal resolution[10] (Extended Data Fig. 6f). We conclude that, under the used illumination conditions, a stable structural state is generated. Although direct evidence that excludes the influence of multiphoton absorption on the observed structural states cannot be readily obtained, it suggests that the observed structural changes are qualitatively correct. The plane of retinal containing the C_{19} methyl, which is located on the opposite side of the isomerizing C_{11}=C_{12} bond, is fixed in the resting state by Thr118$^{3.33}$ (3.36 Å away)[45,46] and Try191^{ECL2}. Consequently, the C_{19} methyl is only minimally affected by the *cis*-to-*trans* isomerization (Fig. 2 and Extended Data Fig. 6), shifting

only half an angstrom towards Tyr191^{ECL2} (with a 36.0° rotation compensated by a backwards elbow movement of the polyene chain) and in the opposite direction to the C_{20} methyl (Fig. 2b,c).

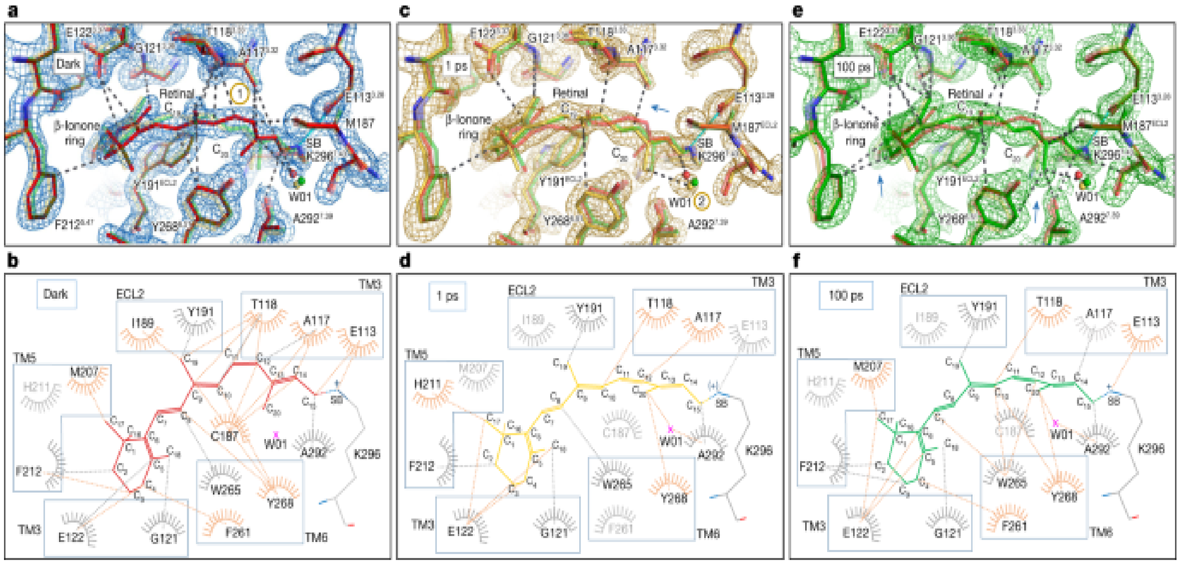
These rearrangements in the retinal molecule at $\Delta t = 1$ ps are compatible with an aborted 'bicycle pedal' mechanism of photoisomerization[14,47,48,49,50] in which the interactions of the C_{19} methyl of retinal with specific residues in the tight binding pocket confer resistance to a larger rotation of this methyl group. The consequence for such aborted C_9=C_{10} isomerization is the release of energy over the polyene chain, affecting mostly C_8 and C_{10}, which elbow in the opposite directions relative to the C_{20} methyl (Fig. 2b,c). The C_6–C_{11} segment of the retinal polyene chain is bent, aligning all carbons in a near-perfect arc (Fig. 2c), affecting the surrounding interactions in the binding pocket.

The absorption maximum of this state was computed by quantum mechanics/molecular mechanics (QM/MM) optimization of the experimental structure at $\Delta t = 1$ ps, and a subsequent vertical excitation energy calculation between the ground and the optically active first excited state of the PSB (Methods). This yielded a spectral red-shift of 32 nm relative to the dark state (for an extended QM system composed of the retinal PSB, Glu113$^{3.28}$, Glu181^{ECL2}, Tyr191^{ECL2}, Tyr268$^{6.51}$, Ser186^{ECL2} and water W01) that is in good agreement with the experimental red-shift of 31 nm (ref. [51]) for Batho-Rh (Extended Data Table 2). Notably, one of the first QM/MM calculations on bovine rhodopsin using the CASPT2//CASSCF/AMBER method on a simple QM system (retinal PSB) yielded a red-shift of about 22 nm (ref. [48]), which is comparable to our calculated value of 24 nm in an equivalent QM system. Our calculations on the extended QM system also show that the twisted all-*trans* retinal at 1 ps stores an excess of 36 kcal mol^{-1} of energy compared with the planar 11-*cis* conformation (for comparison, the energy of a 480-nm photon is 59.6 kcal mol^{-1}), which is in good agreement with the experimental value of 32 kcal mol^{-1} measured for the rhodopsin-to-Batho-Rh transition[52].

Rhodopsin binding pocket at 1 ps

Light-induced isomerization transforms retinal into an agonist that interacts differently with residues in the rhodopsin binding site. One picosecond after photoactivation, the isomerized all-*trans* retinal fills the same volume as the 11-*cis* resting conformation, confirming the hypothesis of space-saving motion[47], but is now free from several hydrogen bonds and van der Waals interactions that stabilize the dark-state structure in an inactive conformation (Fig. 3a,b). The covalently bound retinal bends like an arc stabilized in the middle by steric hindrance and van der Waals interactions between the C_{19} and C_{20} methyl groups of the retinal and two tyrosine residues from the ancestral counterion network, Tyr191^{ECL2} and Tyr268$^{6.51}$. The isomerization and rotation of the C_{20}-methyl–C_{13}–C_{14} plane induce a kink of C_{15} from the polyene chain that influences only minimally the neighbouring PSB/Glu113$^{3.28}$ salt bridge, except the order of the surrounding water W04 in the Glu113 hydrogen bond network (Fig. 2c and Extended Data Fig. 6g,h). The interactions between the polyene chain and Ala117$^{3.32}$–Thr118$^{3.33}$ are two critical retinal contacts with TM3 that are weakened at $\Delta t = 1$ ps (compare Fig. 3a,b with Fig. 3c,d; Extended Data Fig. 7 (top) and Supplementary Video 1). An additional interaction with TM3 is disrupted, between the isomerizing bond and Cys187^{ECL2} of the structurally important and highly conserved disulfide bridge Cys187–Cys110 (ref. [53]), linking TM3 to ECL2 (compare Fig. 3a,b with Fig. 3c,d). Notably, the light-induced structural changes observed around the retinal of bovine rhodopsin bear some topological resemblance to those observed in other low-homology seven-TM-helix retinal-binding proteins from bacteria and archaea. For example, a weakening or disruption of some interactions between retinal and TM3 (which corresponds to helix C in prokaryotic opsins) is also observed (Extended Data Fig. 7 (bottom)) in bacteriorhodopsin[24], the sodium photosensitive pump KR2[24,26] and the chloride pump *Nm*HR[27], which all undergo completely different activation mechanisms. Thus, despite their different evolutionary origins and types of isomerization (11-*cis*-to-all-*trans* versus all-*trans*-to-13-*cis*), retinal-binding proteins seem to share the need to disengage the retinal from the central TM helix before undergoing the next steps of activation.

Fig. 3: Residue environment distances from retinal measured using PyMol and LigPlot software, respectively.

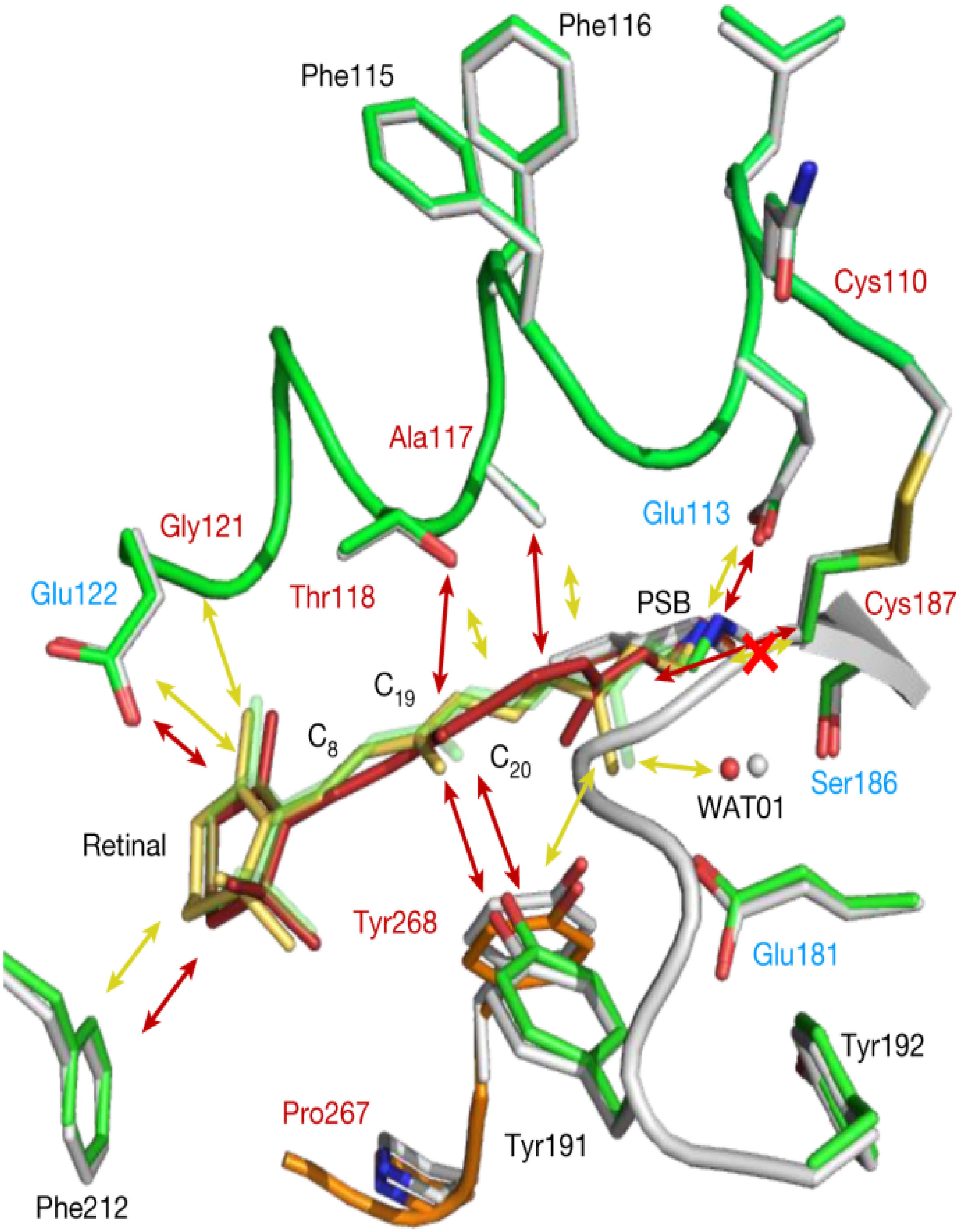


a,**c**,**e**, The two rhodopsin models after 1 ps (yellow) and 100 ps (green) of photoactivation were superimposed in PyMol with the rhodopsin dark-state model (red) and the residue environment distances were drawn with a cut-off at 3.7 Å using dashed lines from the retinal in the dark (**a**) and after 1 ps (**c**) and 100 ps (**e**) photoactivation. The salt bridge between the Schiff base (SB) and counterion Glu113 is marked in cyan. The yellow-circled numbers 1 and 2 show the regions in which interactions will weaken and appear, respectively (see also the blue arrows). **b**,**d**,**f**, While PyMol displays the three-dimensional structure of amino acids in the previous panels, the LigPlot[58] represents a flat interaction plot with all amino acids involved (orange) or less (grey) in the conformational changes during the picosecond time delays of photoactivation (dark (**b**), 1 ps (**d**) and 100 ps (**f**)).

As mammalian rhodopsin evolves along its reaction pathway, some of the observed changes will become more pronounced while others will revert (Fig. 4 and Extended Data Table 3). By $\Delta t = 10$ and 100 ps, the β-ionone ring and C_{19}–C_{20} methyl groups still interact respectively with Gly121$^{3.36}$–Glu122$^{3.37}$ and the tyrosines of the Glu181^{ECL2} polar cage (Fig. 3e,f and Extended Data Fig. 7 (top)), while, for example, Tyr268$^{6.51}$ and Pro267$^{6.50}$ of TM6 relax to their initial positions. Notably, the observed anisotropic character of energy dissipation through TM6 is compatible with intrinsic structural fluctuations in molecular dynamics simulations of rhodopsin in the dark state (Extended Data Fig. 5c,d). This helix is relatively rigid towards

the intracellular G protein-binding site but noticeably more flexible towards the extracellular domains, with the key Pro267$^{6.50}$ at the joint.

Fig. 4: Interactions between retinal and the binding-pocket residues are substantially reduced 1 ps after photoactivation.



Schematic of the interactions of retinal in the rhodopsin ligand-binding pocket before (red arrows) and after photoactivation in the picosecond range (yellow arrows; the red cross on the yellow arrow shows the bond

disruption). A longer arrow represents a stronger interaction. The grey structure corresponds to the dark state (retinal in red) and the coloured structure corresponds to the 1-ps illuminated model (retinal in yellow). For comparison, the retinal model after 100 ps is shown in green. The residues labelled in red are GPCR-conserved and the blue residues are rhodopsin-conserved.

Whereas rhodopsin is considered to be a prototypical class A GPCR, the mechanism used by GPCRs to recognize diffusible agonist ligands by conformational selection is vastly different from the extreme case of induced fit displayed by light-activated GPCRs such as rhodopsin[54]. Notably, both conformational selection and induced fit converge rapidly into a common GPCR-activation mechanism[55] and the early stages of retinal isomerization may therefore reveal fundamental determinants of agonism in GPCRs. For example, early structural changes in the retinal-binding pocket are associated with a small outward tilt (about 0.5 Å) near Pro215$^{5.50}$ and Pro267$^{6.50}$, located around the middle of TM5 and TM6 (Fig. 2b) and enable anisotropic motions in the extracellular part of the receptor (Extended Data Fig. 5). Both proline residues are conserved in class A GPCRs and are key in agonist-induced activation[56]. Moreover, TM3 is a region that has a central role in the architecture of the TM bundle of class A GPCRs[57] by forming part of both the ligand and the G protein-binding pockets. Our TR-SFX data reveal that, even at an early stage of activation, the inverse agonist (11-*cis* retinal) has stripped itself away from TM3 (Figs. 3 and 4 and Supplementary Video 1) while isomerizing into the agonist conformation (Fig. 3c,d). Several of the affected positions in TM3 (Glu113$^{3.28}$, Thr118$^{3.33}$, Gly121$^{3.36}$ and Glu122$^{3.37}$) correspond to conserved residues in the binding site of class A GPCRs that are involved in ligand binding[57], with Gly121$^{3.36}$ in particular being part of a consensus 'cradle' scaffold for ligand recognition[57]. Thus, our TR-SFX observations using XFEL reveal how the photoactivated conformation of the retinal rapidly weakens many van der Waals interactions with the amino acids of the rhodopsin binding pocket (Fig. 4) and thereby commits the receptor's relaxation pathway towards its G protein-binding signalling conformation.

Conclusions

Our high-resolution SFX structure of rhodopsin in the inactive dark state at room temperature reveals the entirety of the water-mediated hydrogen bond network within the protein. One picosecond after light activation, rhodopsin has reached the red-shifted Batho-Rh intermediate. Already by this early stage of activation, the twisted retinal is freed from many of its interactions with the binding pocket while structural perturbations radiate away as a transient anisotropic breathing motion that is almost entirely decayed by 100 ps. Other subtle and transient structural rearrangements within the protein arise in important regions for GPCR activation and bear similarities to those observed by TR-SFX during photoactivation of seven-TM helix retinal-binding proteins from bacteria and archaea. We therefore suggest that the protein disperses an initial excess of energy through the early GPCR structural pathways that will be used for activation. Our study reveals an ultrafast energy dissipation in rhodopsin occurring through conserved residues of GPCR activation pathways and lays the experimental groundwork to study the early activation events in the large family of class A GPCRs.

Methods

Rhodopsin extraction from retinae and purification

All extraction and purification steps were carried out under dim red light conditions. Commercially available dark-adapted frozen bovine retinae (W L Lawson Company) were used to isolate rod outer segment (ROS) membranes according to a protocol described previously[59]. In brief, 200 frozen bovine retinae were diluted in ROS buffer (10 mM MOPS, 30 mM NaCl, 60 mM KCl, 2 mM $MgCl_2$, 1 mM DTT), 40% (w/w) sucrose and two tablets of cOmplete protease inhibitor cocktail. The mixture was shaken by hand for 4 min, centrifuged at 4 °C, 4,000*g* for 30 min. This step was repeated and the pooled supernatants were combined, diluted by a half with ROS buffer containing no sucrose and centrifuged at 4 °C, 24,000*g* for 30 min. The pellets were resuspended in ROS buffer containing 23.4% sucrose and layered onto a freshly prepared gradient with two layers of ROS buffer with 34% (w/w) and 29% (w/w) sucrose. The ROS-membrane-loaded gradients were centrifuged using a swing-out rotor SW28 at 4 °C at 110,000*g* for 90 min and rhodopsin-containing layers (23–29% interface and

29% layer) were aspirated and flash-frozen in liquid nitrogen. The rhodopsin concentration of the dark-state ROS membranes, determined by recording a UV/VIS spectrum before and after illumination, yielded commonly 200 ± 40 mg rhodopsin. Bovine rhodopsin can be further isolated by detergent solubilization and affinity chromatography using a concanavalin A resin (ConA, GE Healthcare Life Sciences) as described previously[60]. This protocol was optimized: (1) the starting retinae material was doubled; (2) the amount of resin was scaled-up three times; (3) to sharpen the elution profile, the second half of the elution phase was performed in reversed flow. A ROS membrane suspension containing around 180 ± 20 mg rhodopsin was diluted three times in ConA buffer (50 mM sodium acetate, 150 mM NaCl, 3 mM $MgCl_2$ $6H_2O$, 3 mM $MnCl_2$ $4H_2O$, 3 mM $CaCl_2$ $2H_2O$, 1 mM Na_2-EDTA $2H_2O$, 2 mM 2-mercaptoethanol, pH 6) and centrifuged at 4 °C, 104,000*g* for 35 min. The resulting ROS membrane pellet was resuspended in 90 ml ConA buffer containing one tablet of protease inhibitor and the membranes were solubilized at room temperature with lauryldimethylamine-oxide (LDAO, Sigma-Aldrich) to a final concentration of 1%. The solubilized sample was centrifuged at 4 °C, 118,000*g* for 60 min before ConA affinity chromatography in ConA buffer containing 0.1% (w/v) LDAO and direct elution with 0.2 M methyl α-d-mannopyranoside in the same buffer. Aliquots of rhodopsin at 2 mg ml^{-1} were flash-frozen in liquid nitrogen and ready to use for crystallization in LCP.

Crystallization and TR-SFX sample preparation

All crystallization and sample preparation steps were performed under dim red light conditions. Before crystallization in LCP, flash-frozen aliquots of purified rhodopsin were thawed and the detergent was exchanged for 0.21% *n*-decyl-*N*,*N*-dimethylamine-*N*-oxide (DAO, Anatrace) in 50 mM sodium acetate, 150 mM NaCl, 3 mM $MgCl_2$, pH 6.0 using a PD10 buffer exchange column (Sigma-Aldrich). The eluate was centrifuged at 18 °C, 21,000*g* for 15 min and further concentrated to at least 20–25 mg ml^{-1} using a concentrator Ultra 4 MWCO 30 at 18 °C and 4,000*g*. After centrifugation at 18 °C and 21,000*g* for 15 min, the protein sample was mixed at 22 °C with monoolein (1-oleoyl-rac-glycerol, Nu-Check prep) at a ratio of 40 to 60, respectively, using gas-tight Hamilton syringes until formation of the translucid LCP. The well diffracting rhodopsin crystal was initially obtained

from a low-molecular-mass polyethylene glycol screen with various types of buffer and the first hit detected using the second harmonic generation imaging SONICC (second order nonlinear optical imaging of chiral crystals) device (Formulatrix)[61]. For crystal growth, 20–40 µl of protein-laden LCP was injected into precipitant-laden Hamilton glass syringes containing 200–400 µl of precipitant (37–39% PEG 600 and 100 mM Bicine pH 9.0). Alternatively, 80 µl of protein-laden LCP were injected into 1 ml plastic syringes containing 800 µl of precipitant. The samples were wrapped in aluminium and stored at 18 °C. After 3 days, plate-shaped crystals in LCP with dimensions of about 15 × 15 × 1.5 µm can be collected by removing the precipitant and kept stable for weeks in the darkness at 18 °C.

At the X-ray free-electron laser beamline, the LCP sample containing rhodopsin crystals was imperatively mixed with a three-way coupler to ensure homogeneity[62] before loading into a reservoir of the high-viscosity injector. In case of residual precipitant contamination, the LCP-laden crystal sample was titrated with PEG 1,000 (50% (w/v)) and finally mixed 1:5 with monoolein.

Time-resolved pump probe serial crystallography

X-ray diffraction data were collected at the XFELs BL3_EH2 end station of the SACLA[63] (beamtimes 2015B8043 and 2018A8066) and Alvra end station of the SwissFEL (beamtimes 20172060 and 20200597). The energy of the X-ray beam was 9–10 keV with a pulse length of 10 fs (SACLA) and 65 fs (SwissFEL) and a focus at the sample of 1 × 1 µm (SACLA) and 5 × 5 µm (SwissFEL). The hutch was prepared for dim red-light conditions and the femtosecond pump laser set at a wavelength of 480 nm with a pulse energy of 9 (SACLA) to 5 (SwissFEL) µJ per 100 fs pulse duration at the sample position. The pump laser beam size was of 47–50 µm FWHM (80–85 µm $1/e^2$). If one assumes idealized Gaussian beam optics, then this corresponds to a peak energy density of 200 mJ cm^{-2} at SwissFEL ($\Delta t = 1$ ps) or, alternatively, a mean energy density of 140 mJ cm^{-2} averaged over the laser's focal FWHM (compare with tabulated data in ref. [3]; or a peak power density of 2,000 GW cm^{-2}). The corresponding values for the SACLA study ($\Delta t = 100$ ps) are a peak fluence of 360 mJ cm^{-2}, and average fluence of 260 mJ cm^{-2}, and a peak power density of 3,600 GW cm^{-2}. The

experimental illumination conditions were chosen to generate a high occupancy of these observed structural states that evolve over time. Although these conditions would be deemed to be excessive in spectroscopic experiments, if one calculates the product of the average fluence (F) with the resting state absorption cross-section (σ) divided by the energy of a single photon (hv, where h is Planck's constant and v the photon's frequency) we recover $\sigma F/hv = 45$ for the illumination conditions at SwissFEL and $\sigma F/hv = 81$ for those used at SACLA. Whereas these values may suggest considerable multiphoton excitation, the results from time-resolved X-ray solution scattering studies on rhodopsin (Extended Data Fig. 3) imply that fewer absorbed photons lead to heating in the sample (see the 'Time-resolved X-ray solution scattering' section below). Moreover, the use of 60% of the energy of the pump laser (3 µJ per 100-fs pulse, 764 GW cm^{-2}) in the 10-ps time delay indicated the lowest energy applicable in our TR-SFX study (Extended Data Fig. 3) and a photoactivation regime higher than single photon per rhodopsin, leaving open the possibility that nonlinear effects contributed to the observed photochemistry.

Crystals of bovine rhodopsin grown in LCP were used to collect TR-SFX data at time delays of 1 ps, 10 ps (SwissFEL) and 100 ps (SACLA) (Extended Data Fig. 2). The crystals were extruded using a high-viscosity injector through a 75-µm nozzle with a constant flow rate of 0.033 µl min^{-1} (SwissFEL) or 2.5 µl min^{-1} (SACLA)[64] to the pump probe intersection point where the data were collected with every fifth shot of the pump laser blocked (data collection scheme of 4 light-activated, then 1 dark) (SwissFEL) or interleaving ON/OFF-laser (collection in the mode 1 light:1 dark (SACLA)), depending on the repetition rates of the XFELs (SwissFEL 25 Hz; SACLA 30 Hz) and pump lasers (SwissFEL 25 Hz; SACLA 15 Hz), respectively. As a control, true dark-state data were also collected with the pump laser off (all dark data, SFX mode).

TR-XSS analysis

TR-XSS studies using samples of detergent-solubilized rhodopsin injected using a gas dynamics virtual nozzle (GDVN) liquid microjet were performed at the LCLS as previously described[65]. Rhodopsin was solubilized in *n*-dodecyl-β-maltoside to a concentration of 8.4 mg ml^{-1} (0.2 mM). The

samples were photoactivated using 480-nm laser pulses 50 fs in duration, focused through a $1/e^2$ spot of 100 µm in diameter (59 µm FWHM) with pulse powers of 6 µJ (110 mJ cm^{-2} averaged across the FWHM; 3,000 GW cm^{-2} peak power), 22 µJ (400 mJ cm^{-2} averaged across the FWHM; 11,200 GW cm^{-2} peak power), 45 µJ (830 mJ cm^{-2} averaged across the FWHM; 22,900 GW cm^{-2} peak power) and 89 µJ (1,640 mJ cm^{-2} averaged across the FWHM; 45,300 GW cm^{-2} peak power). Laser-induced heating was estimated from these data from the time delays of 10 ps $\leq \Delta t \leq$ 1 µs (Extended Data Fig. 3f), but the sparse sampling and signal-to-noise of these data did not allow multiple heating basis spectra to be extracted over this time domain. By contrast, TR-XSS studies on detergent solubilized photosynthetic reaction centre allowed two basis spectra to be extracted over the first 100 ps after photoexcitation. In that study, the amplitude associated with the first heating basis spectrum had reached its maximum value by $\Delta t =$ 10 ps, which then transitioned to a longer-timescale heating basis spectrum with the amplitudes of these two components crossing near $\Delta t = 40$ ps (ref. [65]). The principal singular value decomposition (SVD) component from our rhodopsin TR-XSS data was compared with temperature calibration curves (Extended Data Fig. 3g) to estimate the laser-induced change in temperature for different photoexcitation fluence, as described previously[65]. The heating impulse therefore imparted to the sample is summarized in Extended Data Fig. 3h, with a negative time point used as a control (plotted as zero laser fluence). Whereas these TR-XSS data indicate that laser-induced heating of detergent-solubilized samples of rhodopsin varies linearly with the pump laser pulse fluence, this should not be taken to imply that the photoexcitation is occurring in the single photon per chromophore linear response limit. Moreover, any energy stored within the protein on this time-scale (for example, as strain within the retinal) will not be visible as heating in the TR-XSS data, and therefore any measurable heating above a fraction of a photon equivalent implies that excess energy entered the system through a multiphoton absorption pathway.

Using Extended Data Fig. 3h as a laser-heating-induced calibration curve, we can estimate that the temperature jump induced in detergent solubilized rhodopsin samples injected using a GDVN liquid microjet exposed to the photoexcitation conditions used at SwissFEL (fluence of 140 mJ cm^{-2} averaged over the FWHM) would be $\Delta T = 0.016 \pm 0.009$ °C; and when

exposed to the photoexcitations used at SACLA (fluence of 260 mJ cm^{-2} averaged over the FWHM) would be $\Delta T = 0.028 \pm 0.016$ °C. Using the formula for absorbed photons per molecule = $\Delta T \times C_P$/([rhodopsin] × A × $h\nu$), where $C_P = 3.8$ J cm^{-3} is chosen as an approximate heat capacity for membrane proteins in solution with high detergent content (table 2 of ref. [19]), A is Avogadro's constant, h is Planck's constant, $\nu = c/\lambda$ is the frequency of the pump laser photon, c is the speed of light and λ is the pump laser wavelength. These heating changes would correspond to an excess of 1.2 ± 0.7 photons absorbed by the rhodopsin chromophore under the photoexcitation conditions at SwissFEL, and an excess of 2.1 ± 1.2 photons under the photoexcitation conditions used at SACLA. However, considerable uncertainty must be acknowledged before extrapolating these estimates to our TR-SFX studies. Three of the most important considerations are that appropriate TR-XSS control studies should be performed using similar sample preparations for TR-XSS and TR-SFX studies as well as using the same microjet injector (that is, rhodopsin prepared in LCP and injected using a viscous injector); the above estimates assume that a good spatial overlap between the X-ray beam and laser focus was achieved and this did not drift substantially during the TR-XSS data collection at the LCLS; and ideally the same fs laser should be used for photoexcitation and the same tools should be used to measure the laser focus spot diameter in both TR-XSS and TR-SFX experiments. Despite the additional uncertainty arising from these shortcomings, the above estimates are notably lower than those calculated as $\sigma F/h\nu = 45$ for the photoexcitation conditions used at SwissFEL and $\sigma F/h\nu = 81$ for the photoexcitation conditions use at SACLA ($\sigma_{480\ \mathrm{nm}} = 34{,}000$ M^{-1} cm^{-1}). There has been considerable debate about what constitutes appropriate photoexcitation conditions for TR-SFX studies[3,25,31,32]. Our TR-XSS observations suggest that, as with other TR-XSS studies on other light-sensitive proteins[3], laser-induced sample heating is not accurately predicted by the product $\sigma F/h\nu$. This may be due to the cross-section of the first excited state at 480 nm being much lower than that of the ground state; may be due to the photoexcited states having relatively high stimulated Raman scattering and stimulated emission cross-sections and consequently the absorbed excess energy is carried away by emitted photons rather than being visible as sample heating; there will be some energy loss due to scattering from the microjet; and there may be other

factors that we do not yet understand. After adjusting for rhodopsin having a concentration of 4 mM in crystals but 0.2 mM in the above TR-XSS studies, these heating estimates imply that it is improbable that the temperature jump within crystals was the order of 100 °C as has been claimed for TR-SFX studies of bacteriorhodopsin[32] It is also improbable that our TR-SFX data are dominated by the quasi-isotropic structural expansion due to laser-induced heating that was observed for a photosynthetic reaction centre in TR-XSS studies in which approximately 800 photons were absorbed per chromophore[65].

Data processing

All data were indexed using INDEXAMAJIG with the XGANDALF algorithm for data collected at SwissFEL (SF dark, 1 ps, 10 ps) and the MOSFLM[66], DirAx[67] and XGANDALF[68] algorithms for data collected at SACLA (SACLA dark, 100 ps). The integration radius was set to 2 pixels for SwissFEL data and 3 pixels for SACLA data, while the background annulus was set to between 4 and 6 pixels for SwissFEL data and between 4 and 7 pixels for data collected at SACLA. The crystal-to-detector distance was optimized on a per-run basis by sampling detector distances between 91.5 mm and 97.5 mm (SwissFEL) and between 47.5 mm and 53.5 mm (SACLA) first at 200 µm and then 20-µm increments to determine the detector distance at which the standard deviations of the unit cell dimensions were minimized.

The SwissFEL and SACLA data were scaled and merged separately in PARTIALATOR[69], using partiality modelling with XSPHERE[70]. Custom splitting was used to output separate dark and light-activated reflection files for data collected at each free electron laser.

A lattice translocation defect was identified in the crystals[36] after inspection of the Patterson map with phenix.xtriage[71]. In particular, the Patterson peak at td = (0.000, 0.245, 0.000) (‘Translation vector (*Td*)’ in Extended Data Table 1 and Extended Data Fig. 1d–i) was attributed to the presence of two translation-related domains in the crystals. The correction[36] required to retrieve single-domain intensities was described previously[72], the percentage of molecules in the translated domain (κ) (‘Translated fraction (k)’ in

Extended Data Table 1 and Extended Data Fig. 1d–i) was determined by correcting the intensities at increasing κ values from 0% to 50% in 1% increments until the (0.000, 0.245, 0.000) Patterson peak was flattened. The correction led to a reduction in the SACLA dark state Rfree from 26.11% to 23.92% and improved the interpretability of the dark-state electron density maps, which enabled further improvement of the model (Extended Data Fig. 1d–i).

Structure determination and refinement of rhodopsin dark state

PDB 1U19 (ref. [73]) with solvent and ligand molecules removed was used as a molecular replacement search model in Phaser MR[74]. The dark-state structure was obtained after several iterative cycles of refinement and iterative model building using Phenix.refine[75] and Coot[76]. An additional ligand geometry file was generated using JLigand to restrain the geometry of the PSB linking the lysine side chain to retinal[77].

Calculation of difference density maps

F_o(light) and F_o(dark) amplitudes were calculated from the lattice translation defect corrected intensities using *phenix.french_wilson*[71] and F_o(light) − F_o(dark) difference maps were calculated using phenix.fobs_minus_fobs_map[71] using the multiscaling option excluding amplitudes smaller than 3σ and using reflections within the resolution range between 9 Å and 1.8 Å. All F_{obs}(light) − F_{obs}(dark) were computed using phases of the refined dark state.

For the calculation of $F_{calc} - F_{calc}$ difference maps, the F_{calc} amplitudes were computed using SFall, scaled against experimental data using Scaleit and difference maps were calculated using FFT to a resolution of 1.7 Å, all programs were available in the CCP4 suite[78].

For integrating and plotting density for the fluence response curve, the two 10-ps light datasets and the 1-ps light dataset recorded at SwissFEL were reduced to the size of the smallest dataset (about 29,000 patterns). These data were than scaled and merged as described for the map calculation, and

difference maps were calculated in the exact same way. A custom MATLAB script based on ref. [79] was used to integrate the entire positive difference density in a 2 Å radius around the C_{20} and C_{12} atoms of the excited state. The density was then plotted against the fluence.

Data extrapolation

Extrapolated data were calculated using the lattice translation corrected data and according to a method described previously[80]. A linear approximation was used as follows: $F_{extra} = 100/A \times (F_{ob}(\text{light}) - F_{obs}(\text{dark})) + F_{calc}$, where A is the activation level in percent, F_{extra} represents the extrapolated structure factor amplitudes and F_{calc} represents the calculated amplitudes of the dark-state model. The activation level for each timepoint was determined independently using a previously described method[80]. In brief, extrapolated data were calculated with activation levels ranging from 10% to 50% in 1% increments, and $2F_{extra} - F_{calc}$ difference maps together with phases from the dark-state model were calculated. Negative $2F_{extra} - F_{calc}$ density around C_{11}, C_{12} and C_{20} of retinal, which display negative density features in the $F_{obs}(\text{light}) - F_{obs}(\text{dark})$ maps, was integrated with a radius of 1.5 Å and above at 1.5σ cut-off for each activation level. Negative $2F_{extra} - F_{calc}$ difference density was plotted as a function of activation level; when the activation level is overestimated, there is little $2F_{extra} - F_{calc}$ negative difference density at these atomic positions, while the magnitude of the negative $2F_{extra} - F_{calc}$ difference density increases when the activation level is underestimated. The activation level is then determined by calculating the intersection between the two linear sections of the plot to find the activation level at which the negative density begins to appear. As the three light-activated datasets were collected under different experimental conditions, they were calculated independently (SwissFEL 1 ps 21%, SwissFEL 10 ps 20%, SACLA 100 ps 22%).

Refinement of light-activated states

The dark-state model from SwissFEL was used as an initial model for refinement with the extrapolated data for the 1-ps and 10-ps timepoints in

Phenix.refine[75], interactive model building with Coot[76] was performed to fit the model to the $2F_{\mathrm{extra}} - F_{\mathrm{calc}}$ maps and to remove water molecules lacking electron density. The same iterative procedure was performed starting with the SACLA 100-ps data, using the SACLA dark-state model as a starting point.

Residue numbering

In addition to a number according to their position in the primary sequence, residues in rhodopsin are also assigned a 'general' number according to the Ballesteros–Weinstein scheme[81]. The Ballesteros–Weinstein general number consists of two numbers separated by a dot, where the first denotes the helix (1 to 8) and the second the position relative to the most-conserved residue in that helix, arbitrarily assigned to 50. For example, Glu113$^{3.28}$ denotes that the counterion Glu113 is located in TM3 and twenty-two residues before the most conserved residue in TM3 (Arg135$^{3.50}$).

The Ballesteros numbering for most of the amino acids in this study are Met44$^{1.39}$, Met86$^{2.53}$, Phe91$^{2.58}$, Thr94$^{2.61}$, Glu113$^{3.28}$, Ala117$^{3.32}$, Thr118$^{3.33}$, Cys110$^{3.25}$, Gly120$^{3.35}$, Gly121$^{3.36}$, Glu122$^{3.37}$, Met207$^{5.42}$, His211$^{5.46}$, Phe212$^{5.47}$, Pro215$^{5.50}$, Phe261$^{6.44}$, Trp265$^{6.48}$, Pro267$^{6.50}$, Tyr268$^{6.51}$, Ala269$^{6.52}$, Ala292$^{7.39}$, Phe293$^{7.40}$ and Lys296$^{7.43}$.

Water nomenclature

W01 (chain C/HOH #01) at the tip of C_{20}/RET; W02 (chain C/HOH #02) at Ser186; W03 (chain C/HOH #103) proximal to counterion Glu113 (Gly90, Phe91, Ala117); W04 (chain C/HOH #119) at Met86 (Phe91, Phe116, Ala117), low occupancy increasing by $\Delta t = 1$ ps, resetting by $\Delta t = 100$ ps.

QM/MM calculations

The TR-SFX crystallography structures reported in this work were used as initial geometry for the calculations. The pKa values at pH 9.0 of titratable amino acid residues in the protein were obtained using the PROPKA program[82,83]. Subsequently, the program tleap from the AMBER software

package was used to protonate the protein by considering the previously calculated pKa values[84] We performed first a short (50 steps) molecular mechanics (MM) energy minimization while applying positional restraints to the retinal and Lys296$^{7.43}$ to relieve steric clashes. Subsequently, the geometries of both dark and batho states (1 ps and 100 ps) of rhodopsin were optimized using hybrid QM/MM[85] in the gas phase. The backbone of the protein was kept frozen during the simulation. In the simplest system, the QM part consists of only retinal chromophore and the sidechain of Lys296 that forms the protonated Schiff base (RPSB). The hydrogen link atom (HLA) scheme[86] was used to place the QM/MM boundary in between the C_δ and C_ε atoms of the Lys296 sidechain. We also considered one more extended QM region that includes the proximal counterion (Glu113$^{3.28}$), ancestral counterion (Glu181^{ECL2}), Ser186^{ECL2}, Tyr191^{ECL2}, Tyr268$^{6.51}$ and water W01. The QM part was described using the BP86-D3(BJ) functional[87,88] in conjunction with the cc-pVDZ basis set[89] and the def2/J auxiliary basis set for the resolution of identity[90]. The Chain of Spheres exchange (COSX) algorithm was used in combination with the resolution of identity for the Coulomb term (RI-J). The remaining protein was treated with the Amber ff14SB force field[91]. The TIP3P model was used to describe the water molecules[92]. The QM/MM optimizations were performed by using the quantum chemistry program Orca (v.5.0.2)[93] interfaced with the DL_POLY module of the ChemShell (v.3.7.1) software package[94,95]. The minimized ground-state geometries and partial charges were used to calculate the vertical excitation energies at the RI-ADC(2) level of theory[96] with frozen core orbitals and cc-pVTZ basis set in association with the corresponding auxiliary basis[89]. Moreover, to account for the effect of QM/MM geometry optimization on the excitation, we calculated the vertical excitation energies on the partially MM minimized structures of both dark and batho states for the simplest QM/MM system. The RI-ADC(2) calculations were performed using the Turbomole (v.7.5.1) program package[97]. All of the calculations were performed using the supercomputing facilities at the Paul Scherrer Institute.

Molecular dynamics simulations

We used molecular dynamics simulation data from the GPCRmd database[98], an open access research resource that hosts a comprehensive dataset of molecular dynamics simulations for most GPCR 3D structures solved to date. The GPCRmd offers several tools to analyse simulation trajectories interactively or, alternatively, they can be downloaded and analysed locally. Specifically, we concatenated three simulation replicas (3 × 2,500 frames) of rhodopsin (PDB: 1GZM; trajectory IDs 16414, 16415 and 16416) embedded into a lipid bilayer solvated with water and ions and simulated for an aggregated time of 1.5 µs (that is, 500 ns per replica). We used Python (v.3.10) and the MDanalysis library[99,100] to fetch the simulation data and to compute the average root mean square fluctuation of the protein Cα atoms.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Coordinates and structure factors have been deposited at the PDB under accession codes 7ZBC (rhodopsin in the dark state obtained by SFX at the SACLA), 7ZBE (rhodopsin in the dark state obtained by SFX at the SwissFEL), 8A6C (rhodopsin after 1 ps photoactivation obtained by TR-SFX at the SwissFEL), 8A6D (rhodopsin after 10 ps photoactivation obtained by TR-SFX at the SwissFEL) and 8A6E (rhodopsin after 100 ps photoactivation obtained by TR-SFX at the SACLA).

Code availability

The difference electron density maps ($F_{\mathrm{o(light,\ 1,\ 10\ and\ 100\ ps\ time\ delays)}} - F_{\mathrm{o(dark)}}$) of mammalian rhodopsin photoactivated for 1, 10 and 100 ps, respectively, are available at the Swiss National Supercomputing Center CSCS repository (https://doi.org/10.16907/577a2a74-22ea-4d01-87af-6a99447a430f). The codes for the lattice translation disorder correction and the calculation of the electron density extrapolated maps are available at Zenodo (https://doi.org/10.5281/zenodo.7560364).

References

1. Hofmann, K. P. et al. A G protein-coupled receptor at work: the rhodopsin model. *Trends Biochem. Sci.* **34**, 540–552 (2009).

2. Schoenlein, R. W., Peteanu, L. A., Mathies, R. A. & Shank, C. V. The first step in vision: femtosecond isomerization of rhodopsin. *Science* **254**, 412–415 (1991).

3. Branden, G. & Neutze, R. Advances and challenges in time-resolved macromolecular crystallography. *Science* **373**, eaba0954 (2021).

4. Sakmar, T. P., Franke, R. R. & Khorana, H. G. Glutamic acid-113 serves as the retinylidene Schiff base counterion in bovine rhodopsin. *Proc. Natl Acad. Sci. USA* **86**, 8309–8313 (1989).

5. Palczewski, K. et al. Crystal structure of rhodopsin: a G protein-coupled receptor. *Science* **289**, 739–745 (2000).

6. Terakita, A. et al. Counterion displacement in the molecular evolution of the rhodopsin family. *Nat. Struct. Mol. Biol.* **11**, 284–289 (2004).

7. Kojima, K. et al. Evolutionary steps involving counterion displacement in a tunicate opsin. *Proc. Natl Acad. Sci. USA* **114**, 6028–6033 (2017).

8. Terakita, A., Yamashita, T. & Shichida, Y. Highly conserved glutamic acid in the extracellular IV-V loop in rhodopsins acts as the counterion in retinochrome, a member of the rhodopsin family. *Proc. Natl Acad. Sci. USA* **97**, 14263–14267 (2000).

9. Nakamichi, H. & Okada, T. Local peptide movement in the photoreaction intermediate of rhodopsin. *Proc. Natl Acad. Sci. USA* **103**, 12729–12734 (2006).

10. Nakamichi, H. & Okada, T. Crystallographic analysis of primary visual photochemistry. *Angew. Chem. Int. Ed. Engl.* **45**, 4270–4273 (2006).

11. Standfuss, J. et al. The structural basis of agonist-induced activation in constitutively active rhodopsin. *Nature* **471**, 656–660 (2011).

12. Choe, H. W., Park, J. H., Kim, Y. J. & Ernst, O. P. Transmembrane signaling by GPCRs: insight from rhodopsin and opsin structures. *Neuropharmacology* **60**, 52–57 (2011).

13. Deupi, X. et al. Stabilized G protein binding site in the structure of constitutively active metarhodopsin-II. *Proc. Natl Acad. Sci. USA* **109**, 119–124 (2012).

14. Gozem, S., Luk, H. L., Schapiro, I. & Olivucci, M. Theory and simulation of the ultrafast double-bond isomerization of biological chromophores. *Chem. Rev.* **117**, 13502–13565 (2017).

15. Ernst, O. P. et al. Microbial and animal rhodopsins: structures, functions, and molecular mechanisms. *Chem. Rev.* **114**, 126–163 (2014).

16. Schnedermann, C. et al. Evidence for a vibrational phase-dependent isotope effect on the photochemistry of vision. *Nat. Chem.* **10**, 449–455 (2018).

17. Yang, X. et al. Quantum-classical simulations of rhodopsin reveal excited-state population splitting and its effects on quantum efficiency. *Nat. Chem.* **14**, 441–449 (2022).

18. Moffat, K. The frontiers of time-resolved macromolecular crystallography: movies and chirped X-ray pulses. *Faraday Discuss.* **122**, 65–77 (2003).

19. Barends, T. R. et al. Direct observation of ultrafast collective motions in CO myoglobin upon ligand dissociation. *Science* **350**, 445–450 (2015).

20. Tenboer, J. et al. Time-resolved serial crystallography captures high-resolution intermediates of photoactive yellow protein. *Science* **346**, 1242–1246 (2014).

21. Pande, K. et al. Femtosecond structural dynamics drives the trans/cis isomerization in photoactive yellow protein. *Science* **352**, 725–729 (2016).

22. Coquelle, N. et al. Chromophore twisting in the excited state of a photoswitchable fluorescent protein captured by time-resolved serial femtosecond crystallography. *Nat. Chem.* **10**, 31–37 (2018).

23. Claesson, E. et al. The primary structural photoresponse of phytochrome proteins captured by a femtosecond X-ray laser. *eLife* **9**, e53514 (2020).

24. Nogly, P. et al. Retinal isomerization in bacteriorhodopsin captured by a femtosecond x-ray laser. *Science* **361**, eaat0094 (2018).

25. Nass Kovacs, G. et al. Three-dimensional view of ultrafast dynamics in photoexcited bacteriorhodopsin. *Nat. Commun.* **10**, 3177 (2019).

26. Skopintsev, P. et al. Femtosecond-to-millisecond structural changes in a light-driven sodium pump. *Nature* **583**, 314–318 (2020).

27. Mous, S. et al. Dynamics and mechanism of a light-driven chloride pump. *Science* **375**, 845–851 (2022).

28. Dods, R. et al. Ultrafast structural changes within a photosynthetic reaction centre. *Nature* **589**, 310–314 (2021).

29. Aller, P. & Orville, A. M. Dynamic structural biology experiments at XFEL or synchrotron sources. *Methods Mol. Biol.* **2305**, 203–228 (2021).

30. Wranik, M.et al. Watching the release of a photopharmacological drug from tubulin using time-resolved serial crystallography. *Nat. Commun.* **14**, 903 (2023).

31. Grunbein, M. L. et al. Illumination guidelines for ultrafast pump-probe experiments by serial femtosecond crystallography. *Nat. Methods* **17**, 681–684 (2020).

32. Miller, R. J. D., Pare-Labrosse, O., Sarracini, A. & Besaw, J. E. Three-dimensional view of ultrafast dynamics in photoexcited bacteriorhodopsin in the multiphoton regime and biological relevance. *Nat. Commun.* **11**, 1240 (2020).

33. Barends, T. R. M. et al. Influence of pump laser fluence on ultrafast structural changes in myoglobin. Preprint at https://doi.org/10.1101/2022.11.22.517513 (2022).

34. Caffrey, M. Crystallizing membrane proteins for structure determination: use of lipidic mesophases. *Annu. Rev. Biophys.* **38**, 29–51 (2009).

35. Fotiadis, D. et al. Structure of the rhodopsin dimer: a working model for G-protein-coupled receptors. *Curr. Opin. Struct. Biol.* **16**, 252–259 (2006).

36. Rodrigues, J. R., Casadei C. M., Weinert T., Panneels, V. & Schertler, G. F. X. Correction of rhodopsin serial crystallography diffraction intensities for a lattice-translocation defect. *Acta Crystallogr. D* **79**, 224–233 (2023).

37. Li, J., Edwards, P. C., Burghammer, M., Villa, C. & Schertler, G. F. Structure of bovine rhodopsin in a trigonal crystal form. *J. Mol. Biol.* **343**, 1409–1438 (2004).

38. Yan, E. C. et al. Retinal counterion switch in the photoactivation of the G protein-coupled receptor rhodopsin. *Proc. Natl Acad. Sci. USA* **100**, 9262–9267 (2003).

39. Kimata, N. et al. Retinal orientation and interactions in rhodopsin reveal a two-stage trigger mechanism for activation. *Nat. Commun.* **7**, 12683 (2016).

40. Yoshizawa, T. & Shichida, Y. Low-temperature spectrophotometry of intermediates of rhodopsin. *Methods Enzymol.* **81**, 333–354 (1982).

41. Kukura, P., McCamant, D. W., Yoon, S., Wandschneider, D. B. & Mathies, R. A. Structural observation of the primary isomerization in

vision with femtosecond-stimulated Raman. *Science* **310**, 1006–1009 (2005).

42. Sakmar, T. P. Rhodopsin: a prototypical G protein-coupled receptor. *Prog. Nucleic Acid Res. Mol. Biol.* **59**, 1–34 (1998).

43. Lewis, J. W., Szundi, I., Kazmi, M. A., Sakmar, T. P. & Kliger, D. S. Time-resolved photointermediate changes in rhodopsin glutamic acid 181 mutants. *Biochemistry* **43**, 12614–12621 (2004).

44. Edwards, C. et al. Thermal motion in protein crystals estimated using laser-generated ultrasound and Young's modulus measurements. *Acta Crystallogr. A* **46**, 315–320 (1990).

45. Nagata, T., Oura, T., Terakita, A., Kandori, H. & Shichida, Y. Isomer-specific interaction of the retinal chromophore with threonine-118 in rhodopsin. *J. Phys. Chem.* **106**, 1969–1975 (2002).

46. Yan, E. C. et al. Resonance Raman analysis of the mechanism of energy storage and chromophore distortion in the primary visual photoproduct. *Biochemistry* **43**, 10867–10876 (2004).

47. Warshel, A. Bicycle-pedal model for the first step in the vision process. *Nature* **260**, 678–683 (1976).

48. Andruniow, T., Ferre, N. & Olivucci, M. Structure, initial excited-state relaxation, and energy storage of rhodopsin resolved at the multiconfigurational perturbation theory level. *Proc. Natl Acad. Sci. USA* **101**, 17908–17913 (2004).

49. Frutos, L. M., Andruniow, T., Santoro, F., Ferre, N. & Olivucci, M. Tracking the excited-state time evolution of the visual pigment with multiconfigurational quantum chemistry. *Proc. Natl Acad. Sci. USA* **104**, 7764–7769 (2007).

50. Sen, S., Kar, R. K., Borin, V. A. & Schapiro, I. Insight into the isomerization mechanism of retinal proteins from hybrid quantum mechanics/molecular mechanics simulations. *WIREs Comput. Mol. Sci.* **12**, e1562 (2022).

51. Randall, C. E. et al. A new photolysis intermediate in artificial and native visual pigments. *JACS* **113**, 3473–3485 (1991).

52. Birge, R. R. & Vought, B. W. Energetics of rhodopsin photobleaching: photocalorimetric studies of energy storage in early and later intermediates. *Methods Enzymol.* **315**, 143–163 (2000).

53. Karnik, S. S. & Khorana, H. G. Assembly of functional rhodopsin requires a disulfide bond between cysteine residues 110 and 187. *J. Biol. Chem.* **265**, 17520–17524 (1990).

54. Deupi, X. & Kobilka, B. K. Energy landscapes as a tool to integrate GPCR structure, dynamics, and function. *Physiology* **25**, 293–303 (2010).

55. Venkatakrishnan, A. J. et al. Diverse activation pathways in class A GPCRs converge near the G-protein-coupling region. *Nature* **536**, 484–487 (2016).

56. Hauser, A. S. et al. GPCR activation mechanisms across classes and macro/microscales. *Nat. Struct. Mol. Biol.* **28**, 879–888 (2021).

57. Venkatakrishnan, A. J. et al. Molecular signatures of G-protein-coupled receptors. *Nature* **494**, 185–194 (2013).

58. Wallace, A. C., Laskowski, R. A., Singh, J. & Thornton, J. M. Molecular recognition by proteins: protein-ligand interactions from a structural perspective. *Biochem. Soc. Trans.* **24**, 280–284 (1996).

59. Okada, T., Takeda, K. & Kouyama, T. Highly selective separation of rhodopsin from bovine rod outer segment membranes using combination of divalent cation and alkyl(thio)glucoside. *Photochem. Photobiol.* **67**, 495–499 (1998).

60. Edwards, P. C. et al. Crystals of native and modified bovine rhodopsins and their heavy atom derivatives. *J. Mol. Biol.* **343**, 1439–1450 (2004).

61. Kissick, D. J., Wanapun, D. & Simpson, G. J. Second-order nonlinear optical imaging of chiral crystals. *Ann. Rev. Anal. Chem.* **4**, 419–437

(2011).

62. James, D. et al. Improving high viscosity extrusion of microcrystals for time-resolved serial femtosecond crystallography at X-ray lasers. *J. Vis. Exp.* **144**, e59087 (2019).

63. Tono, K. et al. Diverse application platform for hard X-ray diffraction in SACLA (DAPHNIS): application to serial protein crystallography using an X-ray free-electron laser. *J. Synchrotron Radiat.* **22**, 532–537 (2015).

64. Shimazu, Y. et al. High-viscosity sample-injection device for serial femtosecond crystallography at atmospheric pressure. *J. Appl. Crystallogr.* **52**, 1280–1288 (2019).

65. Arnlund, D. et al. Visualizing a protein quake with time-resolved X-ray scattering at a free-electron laser. *Nat. Methods* **11**, 923–926 (2014).

66. Leslie, A. G. The integration of macromolecular diffraction data. *Acta Crystallogr. D* **62**, 48–57 (2006).

67. Duisenberg, A. Indexing in single-crystal diffractometry with an obstinate list of reflections. *J. Appl. Crystallogr.* **25**, 92–96 (1992).

68. Gevorkov, Y. et al. XGANDALF—extended gradient descent algorithm for lattice finding. *Acta Crystallogr. A* **75**, 694–704 (2019).

69. White, T. A. et al. CrystFEL: a software suite for snapshot serial crystallography. *J. Appl. Cryst.* **45**, 335–334 (2012).

70. White, T. A. Processing serial crystallography data with CrystFEL: a step-by-step guide. *Acta Crystallogr. D* **75**, 219–233 (2019).

71. Adams, P. D. et al. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58**, 1948–1954 (2002).

72. Wang, J., Kamtekar, S., Berman, A. J. & Steitz, T. A. Correction of X-ray intensities from single crystals containing lattice-translocation

defects. *Acta Crystallogr. D* **61**, 67–74 (2005).

73. Okada, T. et al. The retinal conformation and its environment in rhodopsin in light of a new 2.2 Å crystal structure. *J. Mol. Biol.* **342**, 571–583 (2004).

74. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).

75. Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D* **75**, 861–877 (2019).

76. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).

77. Lebedev, A. A. et al. JLigand: a graphical tool for the CCP4 template-restraint library. *Acta Crystallogr. D* **68**, 431–440 (2012).

78. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).

79. Wickstrand, C. et al. A tool for visualizing protein motions in time-resolved crystallography. *Struct. Dyn.* **7**, 024701 (2020).

80. Pandey, S. et al. Time-resolved serial femtosecond crystallography at the European XFEL. *Nat. Methods* **17**, 73–78 (2020).

81. Ballesteros, A. & Weinstein, H. Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. *Methods Neurosci.* **25**, 366–428 (1995).

82. Olsson, M. H., Sondergaard, C. R., Rostkowski, M. & Jensen, J. H. PROPKA3: consistent treatment of internal and surface residues in empirical pKa predictions. *J. Chem. Theory Comput.* **7**, 525–537 (2011).

83. Sondergaard, C. R., Olsson, M. H., Rostkowski, M. & Jensen, J. H. Improved treatment of ligands and coupling effects in empirical calculation and rationalization of pKa values. *J. Chem. Theory Comput.* **7**, 2284–2295 (2011).

84. Case, D. A. et al. *Amber 2021* (Univ. California, San Francisco, 2021).

85. Senn, H. M. & Thiel, W. QM/MM methods for biomolecular systems. *Angew. Chem. Int. Ed. Engl.* **48**, 1198–1229 (2009).

86. Field, M. J., Bash, P. A. & Karplus, M. A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *J. Comput. Chem.* **11**, 700–733 (1990).

87. Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **38**, 3098–3100 (1988)

88. Grimme, S., Ehrlich, S. & Goerigk, L. Effect of the damping function in dispersion corrected density functional theory *J. Comput. Chem.* **32** 1456–1465 (2011).

89. Dunning, T. H. Jr. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **90**, 1007–1023 (1989).

90. Weigend, F. Hartree–Fock exchange fitting basis sets for H to Rn. *J. Comput. Chem.* **29**, 167–175 (2008).

91. Maier, J. A. et al. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **11**, 3696–3713 (2015).

92. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983)

93. Neese, F. The ORCA program system. *WIREs Comput. Mol. Sci.* **2**, 73–78 (2012).

94. Sherwood, P. e al. QUASI: a general purpose implementation of the QM/MM approach and its application to problems in catalysis. *J. Mol. Struct.: THEOCHEM* **632**, 1–28 (2003).

95. Metz, S., Kästner, J., Sokol, A. A., Keal, T. W. & Sherwood, P. ChemShell—a modular software package for QM / MM simulations. *WIREs Comput. Mol. Sci.* **4**, 101–110 (2014).

96. Hättig, C. Structure optimizations for excited states with correlated second-order methods: CC2 and ADC(2). *Adv. Quantum Chem.* **50**, 37–60 (2005).

97. Furche, F. et al. Turbomole. *WIREs Comput. Mol. Sci.* **4**, 91–100 (2014).

98. Rodriguez-Espigares, I. et al. GPCRmd uncovers the dynamics of the 3D-GPCRome. *Nat. Methods* **17**, 777–787 (2020).

99. Gowers, R. et al. MDAnalysis: a Python package for the rapid analysis of molecular dynamics simulations, https://conference.scipy.org/proceedings/scipy2016/oliver_beckstein.html (2019).

100. Michaud-Agrawal, N., Denning, E. J., Woolf, T. B. & Beckstein, O. MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* **32**, 2319–2327 (2011).

101. Aquila, A. et al. Time-resolved protein nanocrystallography using an X-ray free-electron laser. *Opt. Express* **20**, 2706–2716 (2012).

102. Gorel, A., Schlichting, I. & Barends, T. R. M. Discerning best practices in XFEL-based biological crystallography—standards for nonstandard experiments. *IUCrJ* **8**, 532–543 (2021).

103. Neutze, R., Wouts, R., van der Spoel, D., Weckert, E. & Hajdu, J. Potential for biomolecular imaging with femtosecond X-ray pulses. *Nature* **406**, 752–757 (2000).

104. Nakayama, T. A. & Khorana, H. G. Mapping of the amino acids in membrane-embedded helices that interact with the retinal chromophore in bovine rhodopsin. *J. Biol. Chem.* **266**, 4269–4275 (1991).

105. Varma, N. et al. Crystal structure of jumping spider rhodopsin-1 as a light sensitive GPCR. *Proc. Natl Acad. Sci. USA* **116**, 14547–14556 (2019).

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Marie Skłodowska-Curie grant agreements 701647 (to M.J.R.), 701646 (to S. Brünle) and 884104 (PSI-FELLOW-III-3i; to S.S.); the Swiss National Science Foundation through 'Ambizione' grant PZ00P3_174169 (to P.N.) and project grants 31003A_179351 (to J.S.), 192780 (to X.D.) and 310030B_173335 (to G.S.); and through the NCCR:MUST program (to C.J.M. and J.S.). F.P. acknowledges ETH Zürich through the National Center of Competence in Research Molecular Ultrafast Science and Technology and the ETH Femtosecond and Attosecond Science and Technology programs. The project is supported by the Japan Society for the Promotion of Science KAKENHI grant no. 19H05776; the Platform Project for Supporting Drug Discovery and Life Science Research (Basis for Supporting Innovative Drug Discovery and Life Science Research (BINDS)) from AMED under grant no. JP21am0101070 (to S.I.); the National Science Centre (Poland) under grant no. 2017/27/B/ST2/01890 (to A.W.). R.N. acknowledges financial support from the Swedish Research Council (grant no. 2015-00560). We thank the members of the Engineering Team of RIKEN SPring-8 Center for technical support. XFEL experiments were conducted at BL3 of SACLA with the approval of the Japan Synchrotron Radiation Research Institute (proposal numbers 2015B8043 and 2018A8066). We acknowledge the staff at the SwissFEL and the Swiss Light Source synchrotron for excellent performance and support from the Alvra team and Macromolecular Crystallography group, respectively, during the beamtimes allocated for this study. The TR-XSS experiment was performed at the Linac Coherent Light Source (LCLS), SLAC National Accelerator Laboratory. Use of the LCLS is supported by the US Department of Energy, Office of Science, Office of

Basic Energy Sciences under contract no. DE-AC02-76SF00515. Part of the sample injector used at LCLS for this research was funded by the National Institutes of Health, P41GM103393, formerly P41RR001209. We thank T. White and T. Nakane for advice regarding the use of the CrystFEL software and the SACLA processing pipeline, respectively; T. Maier for access to the SONICC imager used for finding the first rhodopsin crystal hits; the members of the High Performance Computing and Emerging Technologies Group at the Paul Scherrer Institute for technical support with the QM/MM simulations; C. Gati for initially helping to measure some rhodopsin crystals controls at the SONICC instrument (Hamburg); and G. Cicchetti for his comments on the paper.

Funding

Open Access funding provided by Lib4RI – Library for the Research Institutes within the ETH Domain: Eawag, Empa, PSI & WSL.

Author information

Author notes

1. Christopher J. Milne

   Present address: European XFEL, Schenefeld, Germany

2. Antonia Furrer

   Present address: Biologics Center, Novartis Institutes for Biomedical Research, Basel, Switzerland

3. Sandra Mous

   Present address: Linac Coherent Light Source, SLAC National Accelerator Laboratory, Menlo Park, CA, USA

4. Petr Skopintsev

Present address: California Institute for Quantitative Biosciences (QB3), University of California, Berkeley, CA, USA

5. Daniel James

   Present address: Department of Physics, Utah Valley University, Orem, UT, USA

6. Steffen Brünle

   Present address: Leiden Institute of Chemistry, Leiden University, Leiden, The Netherlands

7. Filip Pamula

   Present address: Department of Molecular Biology and Genetics, Aarhus University, Aarhus, Denmark

8. Przemyslaw Nogly

   Present address: Dioscuri Center For Structural Dynamics of Receptors, Faculty of Biochemistry, Biophysics and Biotechnology, Jagiellonian University in Kraków, Kraków, Poland

9. These authors contributed equally: Tobias Weinert and Matthew Rodrigues

Authors and Affiliations

1. Division of Biology and Chemistry, Laboratory for Biomolecular Research, Paul Scherrer Institute, Villigen PSI, Switzerland

   Thomas Gruhl, Tobias Weinert, Matthew J. Rodrigues, Antonia Furrer, Petr Skopintsev, Daniel James, Demet Kekilli, Hannah Glover, Steffen Brünle, Azeglio D. Diethelm, Guillaume Gotthard, Pikyee Ma, Jonas Mühle, Filip Pamula, Oliver Tejero, Ching-Ju Tsai, Niranjan Varma, Xavier Deupi, Jörg Standfuss, Gebhard Schertler & Valerie Panneels

2. Photon Science Division, Laboratory for Femtochemistry, Paul Scherrer Institute, Villigen PSI, Switzerland

Christopher J. Milne, Karol Nass, Claudio Cirelli, Camila Bacellar, Dardan Gashi, Victoria Kabanova & Gregor Knopp

3. Department of Chemistry and Molecular Biology, University of Gothenburg, Gothenburg, Sweden

Giorgia Ortolani, Petra Båth, Daniel Sarabi & Richard Neutze

4. Institute of Multidisciplinary Research for Advanced Materials, Tohoku University, Sendai, Japan

Eriko Nango

5. RIKEN SPring-8 Center, Hyogo, Japan

Eriko Nango, Rie Tanaka, Shigeki Owada & So Iwata

6. Condensed Matter Theory Group, Laboratory for Theoretical and Computational Physics, Division of Scientific Computing, Theory and Data, Paul Scherrer Institute, Villigen PSI, Switzerland

Saumik Sen, Ramon Guixà-González & Xavier Deupi

7. Swiss Institute of Bioinformatics (SIB), Lausanne, Switzerland

Saumik Sen, Ramon Guixà-González & Xavier Deupi

8. Photon Science Division, Laboratory for Nonlinear Optics, Paul Scherrer Institute, Villigen PSI, Switzerland

Philip J. M. Johnson

9. Institute of Molecular Biology and Biophysics, Department of Biology, ETH Zurich, Zurich, Switzerland

Sandra Mous, Guillaume Gotthard & Przemyslaw Nogly

10. Photon Science Division, Laboratory for Macromolecules and Bioimaging, Paul Scherrer Institute, Villigen PSI, Switzerland

Florian Dworkowski & Isabelle Martiel

11. Division Scientific Computing, Theory and Data, Paul Scherrer Institute, Villigen PSI, Switzerland

Dmitry Ozerov

12. Department of Cell Biology, Graduate School of Medicine, Kyoto University, Kyoto, Japan

Rie Tanaka & So Iwata

13. Department of Biology, ETH Zurich, Zurich, Switzerland

Cecilia M. Casadei, Elena Lesca & Gebhard Schertler

14. Japan Synchrotron Radiation Research Institute, Hyogo, Japan

Yasumasa Joti, Shigeki Owada & Kensuke Tono

15. Laboratory for Ultrafast X-ray Sciences, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

Victoria Kabanova

16. Institute of Nuclear Physics Polish Academy of Sciences, Kraców, Poland

Anna Wach

17. Operando X-ray Spectroscopy, Energy and Environment Division, Paul Scherrer Institute, Villigen PSI, Switzerland

Anna Wach

18. Linac Coherent Light Source, SLAC National Accelerator Laboratory, Menlo Park, CA, USA

Sébastien Boutet

Contributions

The project was initiated by G.S. with input from R.N. and coordinated by V.P.; G.S. coordinated and supported crystallography applications at SwissFEL and contributed to discussions throughout the project. S.I., E.N. and R.T. shared experience, coordinated and supported crystallography applications at SACLA. Initial XFEL pilot tests at the LCLS and SACLA were supported by S. Boutet, and K.T., Y.J. and S.O., respectively, in a beamtime of J.S. S. Boutet prepared the X-ray instrument and collected the data. Initial rhodopsin crystals were developed in an experiment with T.G., P.M. and V.P. Highly diffracting crystals were optimized and tested for diffraction at the Swiss Light Source by T.G., P.S. and V.P. ROS membranes from bovine retinae were purified by C.-J.T., V.P. and T.G. Rhodopsin was purified and crystallized by T.G., with supplementary support before beamtimes by A.F., S.M., C.-J.T., P.M., J.M., F.P., N.V., W.W. and V.P. Pump–probe experiments at the Alvra endstation of the SwissFEL were prepared by C.J.M., P.J.M.J., C.C. and D.J.; J.S., R.N. and D.J. supported V.P. in setting the laser power and scheme of data collection. The lipidic sample injection was optimized by T.G., N.V. and D.J. Sample preparation and reservoir loading under dark conditions at the XFEL was performed by T.G., A.F., S.M., G.O., D.K., E.L., H.G. and I.M. and injectors for viscous samples were operated during the XFEL beamtimes by G.G., P.B., D.G., O.T. and A.W., and D.J. and F.D., respectively. The endstation, including the laser system, was aligned and operated by P.J.M.J., G.K., C.C., C.B. and C.J.M., who also designed the Alvra prime pump–probe station. The SFX data analysis pipeline of SwissFEL was built and operated by D.O. and K.N. Data processing during the beamtime was performed by T.W., P.N. and K.N. with the help of D.O., B.P., D.K., S. Brünle and C.M.C.; D.K., X.D., C.M.C., A.F. and V.P. recorded progress during data collection. Data processing was further completed by T.W., T.G., K.N., M.J.R., A.D.D. and C.M.C. Crystal lattice translation was identified and corrected by M.J.R., C.M.C., V.P. and T.W. Structures were refined by T.G., M.J.R. and T.W., and data were interpreted by V.P with the help of X.D. and G.S. Quantum mechanics and molecular mechanics calculations were carried out by S.S. Molecular dynamics simulations were analysed by R.G.-G. The TR-XSS

experiment was designed by R.N. and G.S. and analysed by D.S. and R.N. Additional scattering analyses were performed by V.K. The paper was written by V.P. with direct contributions from T.G., X.D., G.S. and R.N., and with further suggestions from the other authors. All of the authors read and acknowledged the manuscript.

Corresponding authors

Correspondence to Gebhard Schertler or Valerie Panneels.

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks Massimo Olivucci and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Rhodopsin molecules crystal packing and lattice translation correction.

a-c.) Rhodopsin molecules packing in the crystal lattice of space group P 2 2_1 2_1. Bovine rhodopsin crystals obtained with the lipidic cubic phase

method reveal a typical molecule packing of type I, consisting of a well-ordered stacking of 2D-crystals **(a.)** The 2D-crystals contact each other through the glycosyl groups of Asn2 and Asn15 of the rhodopsin N-termini, generating head-to-head crystal contacts in the c-dimension. **(b.)** View of the potential physiological dimer[35] contacting the transmembrane domains 1 (in blue) of each of the two rhodopsin molecules. **(c.)** top view of the molecules arrangement. The “Spectrum” rainbow colour transforms gradually from the TM1 in blue to the TM7 and the amphipathic helix 8 in red. **d-i)** Lattice translation correction. Despite a straightforward molecular replacement (Phaser MR, Phenix[75], see Methods) and a solution harbouring a P 2 2_1 2_1 space group, the dark rhodopsin data analysis indicated the presence of more than one off-origin peaks in the Patterson function[36]. In particular, the Patterson peak at td = (0,0.245,0)(SwissFEL) and (0,0.243,0)(SACLA) were attributed to the presence of two translation-related domains within the crystals. Accordingly, a ghost density was identified (**d**) in the Fo-Fc map (in green and red, contoured at 3.83 rmsd) partially overlapping with the 2Fo-Fc map (blue, 2.7 rmsd) from which the rhodopsin model was built in. **e)** After correction of the single domain X-ray intensities[36] according to Wang et al.[72] (see Methods), the overall rhodopsin electron density map 2Fo-Fc displays less ghost electron density (green and red). More locally at a few affected amino acid residues locations, e.g. compare the electron density of P34$^{(1.29)}$/A in the lower panel **(g)** (corrected) to upper **(f)** (original). **h-i)** Importantly, the retinal binding pocket of rhodopsin was not affected by the density overlapping and correction, and the density after correction **(i)** shows only minor changes in the difference map compared to the original data **(h)**.

Extended Data Fig. 2 Time resolved serial femtosecond crystallography.

Time resolved serial femtosecond crystallography (TR-SFX)[101] was conducted using an X-ray free electron laser[102]. Crystal plates (20 µm large and about 1.5 µm thick) were made of rhodopsin purified from bovine retinae (**a. panel**, scale bar is 20 µm). The sample was scaled up (**b. panel**, scale bar is 20 µm) and subjected to a pump and probe experiment (**c. panel**) triggered by a photoactivation at 480 nm (pump laser). Briefly, rhodopsin crystals grown in lipidic cubic phase (LCP) in the darkness were

successively injected (viscous jet) in the light of a pump laser and probed for X-ray diffraction after various time-delays (Δ*t* from 1 to 100 picoseconds) using the X-ray free electron laser from SACLA (Japan)(blue double asterisk) or SwissFEL (Switzerland)(purple asterisk) under the regime “diffract-before-destroy”[103]. The 10 ps time-delay was performed at the SwissFEL with the following laser settings: pump laser, 483 nm, 3 µJ/100fs pulse, 84 µm FWHM and an XFEL at 12 keV with 25 fs pulse length, focus 3 x 5 µm. The processing was done both on the fly and at home using the CrystFEL software[69] combining 30’000 of the diffraction patterns generated by each crystal into a dataset. The inset of the **b. panel** illustrates the opacity of the lipidic phase around the crystal when the rhodopsin crystals are produced in large quantity, hampering spectroscopic experiments on the TR-SFX sample.

Extended Data Fig. 3 Correlation between the power density of the pump laser and either the water heating signal by time-resolved X-ray solution scattering (TR-XSS) or the rhodopsin photoproduct appearance by time-resolved serial femtosecond crystallography.

a-e.) Lowest limit of pump laser power (a-e) for observing decent difference electron density (DED) signals in rhodopsin within the present experimental settings. Two datasets of rhodopsin photoactivated for a 10 ps time delay were collected (SwissFEL beamtime 20200597) at 2 different laser powers, 1.5 (**a and c**) and 3 µJ/100 fs pulse (**b and d**) with the same focus (100 µm spot diameter ($1/e^2$)) and same amount of 29.900 images. For both pump laser powers, the retinal isomerization and the concomitant C20-methyl rotation are observable, but the positive electron density signals are barely detectable in the case of the low power setting (**a panel compared to b**), e.g. the 11-cis-to-trans event is marked by a strong negative electron density on the 11-cis, but no corresponding positive density is detected. When the intensity of the DED is increased for both until an equal level of 0.100 e/Å^3, we observe that the signal-to-noise ratio for the low pump laser energy condition (**c panel**) is so low that the signals cannot be interpreted, compared to the condition with twice the energy (**d panel**). Power titration including 1.5 µJ/100 fs pulse, 3 µJ/100 fs pulse (10 ps time points) and the 5

μJ/100 fs pulse (1 ps time point), with peak power density of 382 GW/cm^2, 764 GW/cm^2 and 1914 GW/cm^2, respectively (**e panel**). **(f-h)** Time-resolved X-ray solution scattering (TR-XSS) studies of visual rhodopsin using XFEL radiation. **f)** TR-XSS difference data (laser on minus laser off) recorded at the LCLS from detergent solubilized samples of rhodopsin for the time-delays 10 ps ≤ Δt ≤ 1 μs at various laser power densities. **g)** Principal singular value decomposition (SVD) component (blue line) from samples of visual rhodopsin indicating laser induced heating (characteristic curve from 0.5 Å^{-1} ≤ q ≤ 2.5 Å^{-1}) as well as oscillations usually associated with protein induced structural changes (visible from 0.25 Å^{-1} ≤ q ≤ 1.0 Å^{-1}). An experimental difference X-ray scattering curve due to heating alone (red line) recorded from detergent solubilized samples of a photosynthetic reaction centre using synchrotron radiation[65], is used to calibrate this laser induced heating. **h**) Laser induced heating in samples of visual rhodopsin measured by TR-XSS when using a 480 nm fs laser pulse with a fluence of 110 mJ/cm^2 (number of independent measurements, n = 6); 400 mJ/cm^2 (n = 6); 830 mJ/cm^2 (n = 3); and 1640 mJ/cm^2 (n = 2), where the fluence is averaged across the FWHM of the laser spot. Negative time-delays (n = 4) were used as a control (plotted at zero fluence). Data are presented as mean values +/− SEM.

Extended Data Fig. 4 Comparison of observed and calculated difference electron density maps at the three time-delays 1, 10 and 100 ps of rhodopsin photoactivation, showing that the refined models (right panels) are in good agreement with raw difference electron density data (left panels).

Difference Fourier electron density maps were created directly from the starting experimental electron density maps (Fobs(light)-Fobs(dark)) **(left panels)** or from the calculated atomic structure model factors (Fcalc(light)-Fcalc(dark)) **(right panels)** and compared for the 1 ps dataset **(a versus b; g versus h)**, 10 ps **(c versus d)**, 100 ps datasets **(e versus f)**. The panels show the retinal binding pocket of rhodopsin in the dark state (**all panels**) (or superimposed with the 1 ps photoactivated structure **(g-f)**) with retinal in red (or in yellow for the 1 ps structure **(g-f)**) and contoured with the 2Fobs-Fcalc electron density map (grey mesh) **(a—f)**. Highlighted in colour, the

difference Fourier electron density signals between photoactivated and dark rhodopsin are displaying features appearing with time, in blue (positive density) that are correlated with disappearing features in gold (negative density). The mesh contouring at various rmsd values was adjusted for easier side to side comparison of the different types of maps and clarity of the figure (see the values in the panels **a—h**).

Extended Data Fig. 5 Anisotropic breathing motion of rhodopsin.

Comparison of the overall conformational changes in rhodopsin photoactivated for 1, 10 and 100 picoseconds. The difference electron density map (Fobs(1ps-light)-Fobs(dark) contoured at 4.2 rmsd) from the dataset of 1ps-illuminated rhodopsin superimposed on the rhodopsin dark state structure model (grey) (**a-b**) shows strong signals (blue=positive density; yellow=negative density) on the retinal molecule (red) demonstrating the early isomerization. Surrounding the retinal, changes occur at the amino acid level in an anisotropic direction towards the extracellular side (grey arrow of **panel a**) along TM5 and TM6 (**panels f and g**). This anisotropic breathing motion can be detected in the extracellular part of TM3 (**e**), TM5 (**f**) and TM6 (**g**). After 10 ps (**panel h**) and 100 ps (**panel i**) of photoactivation, most of the conformational concerted motion changes are reset (**h, i and** Extended Data Table 3), only a few amino acids only will not revert –like the disulphide bridge C110-C187 (pink arrow)- and take part to further changes along the photoactivation pathway (Extended Data Table 3). Interestingly, we also observe localized intrinsic fluctuations in molecular dynamics simulations of rhodopsin in the dark state (PDBid: 1GZM) analysed from the GPCRmd database[98]. These fluctuations localize at the extracellular side of the transmembrane bundle and are compatible with the energy dissipation changes observed at 1 ps (compare **panels c and d (molecular dynamics simulations) with the panels a and b (breathing motion)**). Three independent molecular dynamics simulations (3 x 2500 frames) of rhodopsin with retinal (PDBid: 1GZM) were analysed and the backbone root-mean-square fluctuation (RMSF) was depicted on each residue with increasing values from white to red (the RMSF scale was truncated at 1.8 Å for clarity).

Extended Data Fig. 6 Conformation of retinal after 1, 10 and 100 ps of rhodopsin photoactivation using TR-SFX, and Schiff base surroundings.

(a-f) Retinal conformational changes until 100 ps. **a)** The superimposition of the retinal TR-SFX models in the dark (red model) and 1 to 100 ps photoactivation time delays highlights the main differences: the cis-to-trans isomerization at C11-C12 and the concomitant rotation of the C20-methyl around C13. Beside a slight tilt of the β-ionone ring, another difference between 1 (yellow model), 10 (light blue) and 100 (green) ps-structures is a slight relaxation of the polyene chain towards planarity. **b)** Original electron density map around the retinal in the rhodopsin dark state obtained by SFX (2Fo-Fc map contoured at 2.5 rmsd) and the resulting refined model in red. **c)** Extrapolated electron density map around the retinal of 1 ps-photoactivated rhodopsin obtained by TR-SFX (2Fext-Fc map contoured at 1.9 rmsd) and the resulting refined model in yellow. **d)** Extrapolated electron density map around the retinal of 10ps-photoactivated rhodopsin obtained by TR-SFX (2Fext-Fc map contoured at 0.9 rmsd) and the resulting refined model in blue. **e)** Extrapolated electron density map around the retinal of 100 ps-photoactivated rhodopsin obtained by TR-SFX (2Fext-Fc map contoured at 1.9 rmsd) and the resulting refined model in green. **f)** Structure of rhodopsin after 1ps photoactivation (yellow model) obtained by TR-SFX (this study) compared to a cryo-trapped bathorhodopsin state (green model) and the dark state (red model). **(g-h)** Schiff base-counterion E113 and neighbouring water hydrogen bond network after 1 picosecond of photoactivation. **Panel g**: influence of the C11-C12 isomerization on the Schiff base conformation and distance to the counterion E113. The two models of rhodopsin are superimposed on the Cα atoms of the protein. Retinal after 1 ps of photoactivation (yellow (with orange K296[(7.43)])) is showing an all-trans conformation and the C15 of the C14-C15-NZ plane at the SB displays a slight kick towards the extracellular space compared to the structure of the dark state (red). The counterion E113[(3.28)] moves accordingly, in the same direction of about 0.2-0.3 Å. **Panel h**: of the two water molecules W03 and W04 which form a bridge between the counterion E113[(3.28)] and M86[(2.53)] (and also contacting A117[(3.32)], F91[(2.58)] and F116[(3.31)], not shown), only W04 has gained order. The two rhodopsin molecules models (dark in red; 1 ps in yellow) are contoured with their

respective electron density maps, in blue (2Fo-Fc contoured at 1.3 rmsd) and in orange (2Fextrapolated-Fc contoured at 1.3 rmsd). By $\Delta t = 100$ ps we observe a reset of the occupancy, which is similar to that of the dark state structure.

Extended Data Fig. 7 Interactions of retinal with its binding pocket at 0, 1, 10 and 100 ps of photoactivation and comparison with microbial rhodopsins.

LIGPLOT[58] view of retinal interactions within the rhodopsin binding site **(upper four panels)** (set distance < 3.6 Å) at different time-delays of photoactivation, dark state, 1, 10 and 100 ps-photoactivated states. The amino acids and dashed lines of interatomic interactions labelled in orange are the site of major changes, showing new interactions, e.g. with water W01 or losing contact with C187[(ECL2)]. Retinal binding pocket of non-homologous rhodopsins **(lower four panels)** (PDBid codes shown in bold) from mammalian (the rhodopsin GPCR from Bos taurus (this study) (PDBid=**7ZBE**)) and prokaryotes (bacteriorhodopsin proton pump from Halobacterium salinarum (PDBid=**6G7H**); KR2 sodium pump from *Krokinobacter eikastus* (PDBid=**6TK7**) and the NmHR chloride pump from *Nonlabens marinus* (PDBid=**7O8F**)). The third transmembrane helix (“TM3” in GPCRs and “helix C” in prokaryotes) has been described as a main interaction site for retinal[104], often carrying the stabilizing counterion, like E113 in rhodopsin. Projected on the dark state structure of these four rhodopsins, some important structural rearrangements upon retinal isomerization (*cis*-to-*trans* for mammalian rhodopsin, *trans*-to-*cis* for prokaryotic rhodopsins), observed in TR-SFX studies in the picosecond range, are indicated with a red arrow. The disruption of these important interactions weakens the stabilization of the retinal chromophore by TM3/helix C.

Extended Data Table 1 Crystallographic data statistics for rhodopsin structures after 1, 10 and 100 picoseconds of photoactivation, compared to the dark state

Extended Data Table 2 Vertical excitation energies computed by quantum mechanics/molecular mechanics (QM/MM) calculation

Extended Data Table 3 Amplitude of amino acids shifts (Å) after 1ps photoactivation

Supplementary information

Reporting Summary

Supplementary Video 1

Time-lapse GIF image of the retinal-binding pocket after 0 ps, 1 ps and 100 ps of rhodopsin photoactivation. The three models of rhodopsin in the dark (red) and at 1 ps (yellow) and 100 ps (green) of photoactivation are superimposed. The video shows a time-lapse GIF of the retinal-binding pocket of those models contoured successively with the electron density of the dark ($2F_o - F_c$), then the 1 ps ($2F_{ext} - F_c$), then 100 ps ($2F_{ext} - F_c$) maps. The many weak interactions are shown as dashed lines, and yellow arrows point to the regions changing the most.

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Gruhl, T., Weinert, T., Rodrigues, M.J. *et al.* Ultrafast structural changes direct the first molecular events of vision. *Nature* **615**, 939–944 (2023). https://doi.org/10.1038/s41586-023-05863-6

- Received: 02 April 2022
- Accepted: 17 February 2023
- Published: 22 March 2023
- Issue Date: 30 March 2023
- DOI: https://doi.org/10.1038/s41586-023-05863-6

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

This article is cited by

- **Earliest molecular events of vision revealed**
    - Marius Schmidt
    - Emina A. Stojković

  Nature (2023)

Earliest molecular events of vision revealed

- Marius Schmidt
- Emina A. Stojković

Nature News & Views 22 Mar 2023

This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05863-6

| Section menu | Main menu |

Amendments & Corrections

- **Author Correction: An E1–E2 fusion protein primes antiviral immune signalling in bacteria** [13 Mar 2023]
  -
- **Publisher Correction: Synaptic gradients transform object location to action** [13 Mar 2023]
  -
- **Publisher Correction: Essential elements of radical pair magnetosensitivity in Drosophila** [15 Mar 2023]
  -
- **Publisher Correction: A condensate dynamic instability orchestrates actomyosin cortex activation** [16 Mar 2023]
  -

- Author Correction
- Published: 13 March 2023

Author Correction: An E1–E2 fusion protein primes antiviral immune signalling in bacteria

- Hannah E. Ledvina,
- Qiaozhen Ye,
- Yajie Gu,
- Ashley E. Sullivan,
- Yun Quan,
- Rebecca K. Lau,
- Huilin Zhou,
- Kevin D. Corbett &
- …
- Aaron T. Whiteley

Show authors

Nature **volume 615**, page E25 (2023)

- 410 Accesses
- 2 Altmetric
- Metrics details

The Original Article was published on 08 February 2023

This article has been updated

Correction to: *Nature* https://doi.org/10.1038/s41586-022-05647-4
Published online 8 February 2023

In the version of this article originally published, the upper band in Figure 3d labelled 65 kDa was mistakenly taken from the 35 kDa band (see Supplementary Fig. 1 below). The correct 65 kDa selection now appears in the HTML and PDF versions of the figure.

Supplementary information is available in the online version of this notice.

Change history

- **16 March 2023**

  This amendment was updated to include the Supplementary Information file originally published with the article.

Author information

Author notes

1. These authors contributed equally: Hannah E. Ledvina, Qiaozhen Ye

Authors and Affiliations

1. Department of Biochemistry, University of Colorado Boulder, Boulder, CO, USA

   Hannah E. Ledvina, Ashley E. Sullivan & Aaron T. Whiteley

2. Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, CA, USA

Qiaozhen Ye, Yajie Gu, Yun Quan, Rebecca K. Lau, Huilin Zhou & Kevin D. Corbett

3. Biomedical Sciences Graduate Program, University of California, San Diego, La Jolla, CA, USA

Rebecca K. Lau

Corresponding authors

Correspondence to Kevin D. Corbett or Aaron T. Whiteley.

Supplementary information

Supplementary Fig. 1

This file contains the original uncut gels.

Rights and permissions

Reprints and Permissions

About this article

Cite this article

Ledvina, H.E., Ye, Q., Gu, Y. *et al.* Author Correction: An E1–E2 fusion protein primes antiviral immune signalling in bacteria. *Nature* **615**, E25 (2023). https://doi.org/10.1038/s41586-023-05931-x

- Published: 13 March 2023
- Issue Date: 30 March 2023
- DOI: https://doi.org/10.1038/s41586-023-05931-x

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05931-x

- Publisher Correction
- Open Access
- Published: 13 March 2023

Publisher Correction: Synaptic gradients transform object location to action

- Mark Dombrovski,
- Martin Y. Peek,
- Jin-Yong Park,
- Andrea Vaccari,
- Marissa Sumathipala,
- Carmen Morrow,
- Patrick Breads,
- Arthur Zhao,
- Yerbol Z. Kurmangaliyev,
- Piero Sanfilippo,
- Aadil Rehan,
- Jason Polsky,
- Shada Alghailani,
- Emily Tenshaw,
- Shigehiro Namiki,
- S. Lawrence Zipursky &
- …
- Gwyneth M. Card

Show authors

Nature **volume 615**, page E26 (2023)

- 311 Accesses

- 2 Altmetric

- Metrics details

The Original Article was published on 04 January 2023

Correction to: *Nature* https://doi.org/10.1038/s41586-022-05562-8
Published online 4 January 2023

In the version of this article originally published, the *x*-axis labels for Figure 3e,f,i and j incorrectly read “Disc size (°)” and have been corrected to read “Azimuth (°)” in the HTML and PDF versions of the article.

Author information

Author notes

1. Shigehiro Namiki

   Present address: Research Center for Advanced Science and Technology, University of Tokyo, Tokyo, Japan

2. Gwyneth M. Card

   Present address: Department of Neuroscience, Howard Hughes Medical Institute, The Mortimer B. Zuckerman Mind Brain Behavior Institute, Columbia University, New York, NY, USA

3. These authors contributed equally: Mark Dombrovski, Martin Y. Peek, Jin-Yong Park

Authors and Affiliations

1. Department of Biological Chemistry, Howard Hughes Medical Institute, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA

Mark Dombrovski, Yerbol Z. Kurmangaliyev, Piero Sanfilippo, Aadil Rehan & S. Lawrence Zipursky

2. Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA, USA

   Martin Y. Peek, Jin-Yong Park, Marissa Sumathipala, Carmen Morrow, Patrick Breads, Arthur Zhao, Jason Polsky, Shada Alghailani, Emily Tenshaw, Shigehiro Namiki & Gwyneth M. Card

3. Department of Computer Science, Middlebury College, Middlebury, VT, USA

   Andrea Vaccari

Corresponding authors

Correspondence to S. Lawrence Zipursky or Gwyneth M. Card.

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Dombrovski, M., Peek, M.Y., Park, JY. *et al.* Publisher Correction: Synaptic gradients transform object location to action. *Nature* **615**, E26 (2023). https://doi.org/10.1038/s41586-023-05930-y

- Published: 13 March 2023
- Issue Date: 30 March 2023
- DOI: https://doi.org/10.1038/s41586-023-05930-y

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05930-y

- Publisher Correction
- Open Access
- Published: 15 March 2023

Publisher Correction: Essential elements of radical pair magnetosensitivity in *Drosophila*

- Adam A. Bradlaugh,
- Giorgio Fedele,
- Anna L. Munro,
- Celia Napier Hansen,
- John M. Hares,
- Sanjai Patel,
- Charalambos P. Kyriacou,
- Alex R. Jones,
- Ezio Rosato &
- …
- Richard A. Baines

Show authors

Nature **volume 615**, page E27 (2023)

- 377 Accesses
- 1 Altmetric
- Metrics details

The Original Article was published on 22 February 2023

Correction to: *Nature* https://doi.org/10.1038/s41586-023-05735-z
Published online 22 February 2023

In the version of this article initially published, the Author contributions section was truncated. The first two sentences have been corrected to read "R.A.B., A.A.B., G.F., A.R.J., C.P.K. and E.R. conceived the project. A.A.B., G.F., A.L.M. and C.N.H. performed experiments" in the HTML and PDF versions of the article.

Author information

Authors and Affiliations

1. Division of Neuroscience, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester Academic Health Science Centre, Manchester, UK

   Adam A. Bradlaugh, Anna L. Munro & Richard A. Baines

2. Department of Genetics and Genome Biology, University of Leicester, Leicester, UK

   Giorgio Fedele, Celia Napier Hansen, John M. Hares, Charalambos P. Kyriacou & Ezio Rosato

3. Department of Biology, Ecology and Earth Sciences (DiBEST), University of Calabria, Rende, Italy

   Giorgio Fedele

4. Pelican Healthcare, Cardiff, UK

   John M. Hares

5. Manchester Fly Facility, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK

   Sanjai Patel

6. Biometrology, Chemical and Biological Sciences Department, National Physical Laboratory, Teddington, UK

   Alex R. Jones

Corresponding authors

Correspondence to Ezio Rosato or Richard A. Baines.

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Bradlaugh, A.A., Fedele, G., Munro, A.L. *et al.* Publisher Correction: Essential elements of radical pair magnetosensitivity in *Drosophila*. *Nature* **615**, E27 (2023). https://doi.org/10.1038/s41586-023-05929-5

- Published: 15 March 2023

- Issue Date: 30 March 2023
- DOI: https://doi.org/10.1038/s41586-023-05929-5

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05929-5

- Publisher Correction
- Open Access
- Published: 16 March 2023

Publisher Correction: A condensate dynamic instability orchestrates actomyosin cortex activation

- Victoria Tianjing Yan,
- Arjun Narayanan ,
- Tina Wiegand,
- Frank Jülicher &
- …
- Stephan W. Grill

Show authors

Nature **volume 615**, page E28 (2023)

- 404 Accesses
- 2 Altmetric
- Metrics details

The Original Article was published on 17 August 2022

Correction to: *Nature* https://doi.org/10.1038/s41586-022-05084-3
Published online 17 August 2022

In the version of this article initially published, in the legend for Fig. 3c, WSP-1 was erroneously denoted as magenta and F-actin as green. The corrected text now reads “Thick lines indicate the measured WSP-1 (green) and F-actin (magenta) nullclines”. In addition, in the x-axis label of Fig. 4d, “Lifeact” erroneously appeared as “Lifetime”. The corrected label now reads “Lifeact concentration (IU μm^{-3})”. The errors have been corrected in the HTML and PDF versions of the article.

Author information

Author notes

1. These authors contributed equally: Victoria Tianjing Yan, Arjun Narayanan

Authors and Affiliations

1. Max Planck Institute of Molecular Cell Biology and Genetics (MPI-CBG), Dresden, Germany

   Victoria Tianjing Yan, Arjun Narayanan, Tina Wiegand & Stephan W. Grill

2. Biotechnology Center, TU Dresden, Dresden, Germany

   Victoria Tianjing Yan & Arjun Narayanan

3. Max Planck Institute for the Physics of Complex Systems (MPI-PKS), Dresden, Germany

   Arjun Narayanan, Tina Wiegand & Frank Jülicher

4. Center for Systems Biology Dresden (CSBD), Dresden, Germany

   Arjun Narayanan, Tina Wiegand, Frank Jülicher & Stephan W. Grill

5. Cluster of Excellence Physics of Life, TU Dresden, Dresden, Germany

Frank Jülicher & Stephan W. Grill

Corresponding authors

Correspondence to Arjun Narayanan, Frank Jülicher or Stephan W. Grill.

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Yan, V.T., Narayanan, A., Wiegand, T. *et al.* Publisher Correction: A condensate dynamic instability orchestrates actomyosin cortex activation. *Nature* **615**, E28 (2023). https://doi.org/10.1038/s41586-023-05932-w

- Published: 16 March 2023
- Issue Date: 30 March 2023
- DOI: https://doi.org/10.1038/s41586-023-05932-w

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05932-w

Nature Outline

- **Acute kidney injury** [29 Mar 2023]
  • Acute kidney injury, a common occurrence among hospital patients, can trigger a range of long-term health problems.

| Main menu | Previous section |

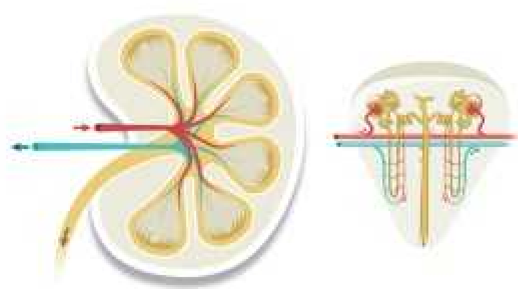


Preventing kidney injuries in hospital

Delayed diagnosis and limited treatment options leave people with acute kidney injury at risk of long-term health problems. Researchers are now looking for ways to act earlier and more effectively.

- Michael Eisenstein

Outline 29 Mar 2023 Nature

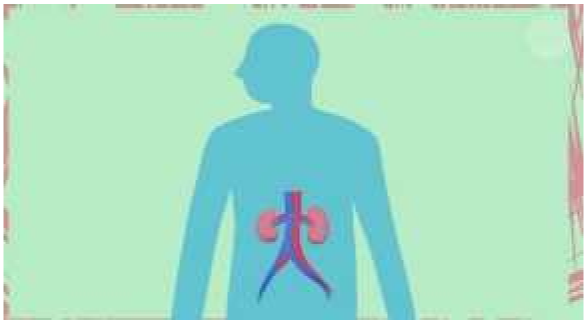


What is acute kidney injury? A visual guide

Early detection and prompt treatment could prevent long-term health effects of acute kidney injury, a condition that commonly arises while people are in hospital.

- Michael Eisenstein

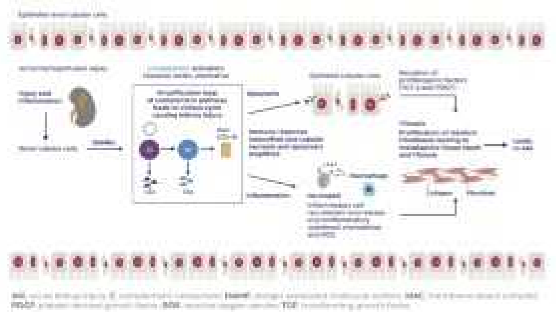
Outline 29 Mar 2023 Nature



Video: Safeguarding the kidney

Physicians aim to find better ways to prevent, diagnose and treat acute kidney injury.

Outline 29 Mar 2023 Nature



The potential role of complement in cardiac surgery-associated acute kidney injury

Sponsor Feature 29 Mar 2023

This article was downloaded by **calibre** from https://www.nature.com/collections/biejeacehj